



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES "ACATLAN"

DISEÑO DE UN SISTEMA DE INFORMACION ESTRATEGICO ADQUIRENTE PARA PRS

MEMORIAS DE DESEMPEÑO QUE PARA OBTENER EL TITULO DE: LICENCIADO EN MATEMATICAS APLICADAS Y COMPUTACION PRESENTA: NAYELLI YURIDIA VARGAS TORICES

ASESOR: JOSE ALFREDO LOPEZ RODRIGUEZ

278453



CAMPUS ACATLÁN MEXICO, D. F.

ABRIL DEL 2000.



Universidad Nacional  
Autónoma de México



**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.

A DIOS, por acompañarme en cada momento de mi vida llenándome de su eterno amor, de su sabiduría, inteligencia e infinita bondad lo que me permite comprender y amar a los demás

A mis papás Gloria y Juan, por brindarme en cada instante su apoyo incondicional, por creer siempre en mí, por haberme enseñado los más elevados principios y valores, y sobre todo, por entregarme lo mejor que hay en la vida, su amor.  
A ustedes todo mi amor y eterno agradecimiento

A mis hermanos y amigos, por haber estado a mi lado en los momentos buenos y malos, por su cariño y sus enseñanzas

A Vadi (Azul), quien me ha enseñado la belleza de la magia y la grandiosidad del amor



## INDICE

Capítulo I.	Introducción	3
Capítulo II.	Información generalcerca de la organización.	6
	Historia	7
	Presente	9
	Esquema general de la empresa	9
	Papel del sistema de información estratégica en la empresa	11
	Reflexión de la información general acerca de la organización	16
Capítulo III.	Análisis de requerimientos de la empresa	17
	Reflexión del análisis de requerimientos	24
Capítulo IV.	OLAP (On Line Analytical Processing)	26
	Características de OLAP	29
	Base de datos relacionales OLAP (ROLAP)	34
	Base de datos multidimensionales OLAP (MOLAP)	34
	Consejos al crear un MDDB	37
	Reflexión de OLAP (On Line Analytical Processing )	40
Capítulo V.	Data Warehouse (DWH)	44
	Introducción	45
	Definiciones de Data Warehouse	45
	Características técnicas de un DWH	47
	Pasos para la creación de un DWH	48
	Reflexión del Data Warehouse	63
Capítulo VI.	Herramienta SAS	65
	Historia	66
	Base de datos multidimensional SAS	73
	HOLAP (Hybrid On line Analitical Processing)	79
	Arquitectura HOLAP	81



Reflexión de la herramienta SAS	88
Capítulo VII. Diseño del sistema	90
Lista de eventos	92
Diagrama de Contexto.	92
Diagrama de Flujo de Datos (DFD).	93
ANÁLISIS DE TABLAS	100
Capítulo VIII. Conclusiones	106
BIBLIOGRAFIA	109
ANEXO	111



# Capítulo I.

## Introducción

**¿Cómo puede ser que las matemáticas, que son después de todo un producto del pensamiento humano, independiente de la experiencia, se adapten tan admirablemente a los objetos de la realidad?**

**Albert Einstein**



El manejo apropiado y oportuno de toda la información al alcance de las entidades que deben tomar decisiones dentro de las organizaciones se ha convertido en el arma más efectiva para la realización de negocios exitosos a fines del siglo XX.

Hoy en día se ha vuelto indispensable contar con repositorios de información los cuales sean consistentes y completos, estos repositorios se vuelven necesarios ya que deben contar con información de las operaciones de la empresa. Así mismo, se ha requerido recurrir a herramientas que puedan permitir la explotación de la información, es por ello que ahora muchas herramientas se enfocan en esto y en la presentación de información.

Inicialmente las empresas se preocupaban por la automatización de los procesos, sin importar tanto la información que se produjera. Conforme van creciendo las empresas, va aumentando la información que se genera, ya que se tienen mas procesos. Importa por tanto tener información confiable que represente la tarea que realiza la empresa, ya que ahora el análisis de dichos datos es primordial para la empresa para realizar estrategias de mercado, producción, financiamiento y actualización tecnológica. Se debe tener en cuenta que hablar de toda la información que encierran los sistemas es demasiada y probablemente repetitiva, además de que puede también ser inconsistente. De aquí que nazca el concepto del Data Warehouse, con el cual se puede tener información necesaria e importante de cierto(s) tema(s).

El presente trabajo da un ejemplo de esto. Se trata la problemática de una empresa con la necesidad de almacenar información confiable para la toma de decisiones y poder presentarla a los clientes. Así mismo se enuncian los conceptos OLAP (On Line Analytical Process) que se encuentran muy relacionados con el DWH y la manera en que este se puede explotar.

Una vez que se tienen información confiable se requiere presentarla al usuario



(cliente), para ello es necesario contar con herramientas que faciliten la presentación de esta, además de que cuenten con un formato amigable y fácil de manejar. En el caso de este trabajo se utilizó tanto para la presentación de información, como la generación de bases de datos la herramienta SAS.





# Capítulo II.

## Información general acerca de la organización.

**Lleva menos tiempo hacer algo bien que  
explicar porque se hizo mal**

**Henry Wadsworth Longfellow**



## Historia

En 1968, un grupo de 10 bancos se unieron para crear en conjunto la infraestructura de su propia tarjeta de crédito<sup>1</sup> y así competir agresivamente con los sistemas de pago de esa época.

De esta unión nace una empresa financiera dedicada a la operación de tarjetas de crédito, desde la propia emisión del plástico hasta el control de facturación.

En 1988 entra en vigor el sistema RED Cajeros Automáticos, logrando que dicha empresa tenga la capacidad de prestar servicios de banca electrónica compartida a los bancos asociados.

Más tarde, se da una expansión en las actividades de la empresa y es entonces cuando se convierte en miembro del sistema VISA<sup>2</sup>.

En 1993, en respuesta a la privatización de la banca y a la apertura de las fronteras, la compañía se asocia con Total System de Estados Unidos, compañía procesadora de transacciones bancarias líder en América quien se encarga de toda la función de emisor<sup>3</sup>, es decir, todo aquello relacionado con la emisión de las tarjetas de crédito y el control de la información y los estados de cuenta.

Con el tiempo, algunas funciones fueron transferidas a los bancos y otras a prestadores externos de servicios. De esta forma la empresa se convierte en un centro de utilidades para los bancos, es decir, empieza a prestar servicios específicos a los bancos del consorcio, principalmente orientados hacia la

---

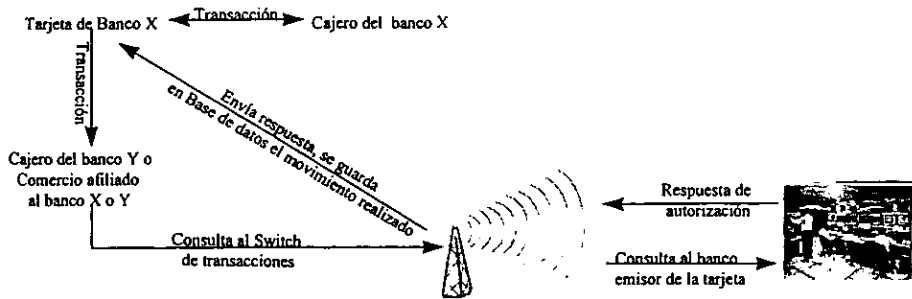
<sup>1</sup> Instrumento mediante el cual una persona puede adquirir bienes o servicios y dispone de cierto plazo para pagarlos, o bien utiliza el financiamiento que el banco le ofrece pagando los intereses correspondientes.

<sup>2</sup> Sistema de tarjetas de crédito que permite que éstas sean reconocidas mundialmente.

<sup>3</sup> Banco que expide las tarjetas y maneja la cuenta del tarjetahabiente.



compensación y liquidación<sup>4</sup> de las transacciones, así como a la especialización del switch de transacciones.



Funcionamiento del switch de transacciones

Como puede observarse en el diagrama anterior, cuando un tarjetahabiente genera una transacción sea esta en un comercio afiliado a un banco del consorcio (sólo cuando la transacción es realizada en terminales punto de venta (POS<sup>5</sup>)) o un cajero al que no pertenece la tarjeta (a estos dos los llamaremos establecimientos), la transacción se envía al switch, el cual hace el papel de intermediario entre establecimiento y banco al que corresponde la tarjeta. Al realizar la transacción, el switch se conecta al banco que corresponde la tarjeta pidiendo autorización para la operación que se desea realizar. El banco envía una respuesta y el switch la regresa al establecimiento. En este proceso se guarda información de la operación realizada, tal como:

- El número de cuenta del tarjetahabiente.
- Sucursal.
- Importe de la transacción.

<sup>4</sup> Proceso que se da entre bancos, mediante el cual se saldan las operaciones realizadas durante el día.

<sup>5</sup> Dispositivo electrónico capaz de procesar transacciones de venta y/o devolución en un comercio con tarjeta de crédito o débito y proporcionar reportes que faciliten el control en sus transacciones y en algunos casos la administración de su negocio.



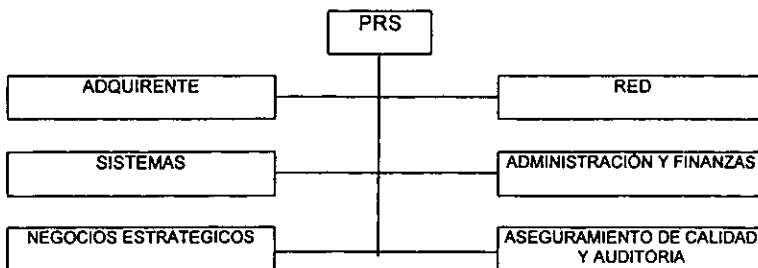
- Banco emisor de la tarjeta.
- Banco dueño del cajero o al que se encuentra afiliado el comercio.
- Tipo de transacción (crédito o débito).
- Fecha en que se realizó la transacción.
- Clave del comercio o del banco.

## Presente

En la actualidad, la empresa se encarga de procesar tarjetas de crédito, débito, operar cajeros automáticos y puntos de venta, así como compensar y liquidar localmente las transacciones generados por los bancos.

Con base en los cambios de los últimos años, la visión de la empresa se enfoca a ser un aliado estratégico altamente reconocido en el mercado de medios de pago para los bancos e instituciones participantes, ofreciendo a sus clientes los servicios y productos especializados afines al switch de transacciones electrónicas, compensación, liquidación y servicios asociados.

## Esquema general de la empresa



Organigrama general de la empresa



A continuación se hace una descripción de cada una de las áreas (conocidas como áreas de negocio) en que se encuentra dividida la empresa:

### **Adquirente**

Se encarga de operación de negocios Adquirente<sup>6</sup>, que consiste en la integración de los comercios al sistema de tarjetas de la empresa con el fin de que estos puedan llevar a cabo operaciones principalmente en terminales POS y en medios de pago en general, agilizando y facilitando el procesamiento de dichas operaciones en el ámbito bancario y comercial.

### **Red**

Se encarga del procesamiento de transacciones electrónicas, fungiendo como enlace entre las instituciones financieras mexicanas y redes nacionales e internacionales en lo que se refiere a cajeros automáticos.

### **Sistemas**

Evalúa, diseña y opera la plataforma tecnológica requerida para proporcionar los servicios a las unidades del negocio.

### **Administración y finanzas**

Coordina, integra y desarrolla al personal que labora en la empresa y asegura el equilibrio de los estados financieros de tal negocio.

---

<sup>6</sup> Banco que incorpora un comercio y le proporciona un servicio.



## **Negocios estratégicos y mercadotecnia**

Identifica, desarrolla y comercializa nuevos productos que sean de valor agregado para los clientes de la empresa.

## **Aseguramiento de calidad y auditoría**

Coordina y desarrolla la cultura y el sistema de aseguramiento de calidad de la compañía, así como la medición del desempeño de los procesos y determina la magnitud y naturaleza de los riesgos que enfrenta la empresa.

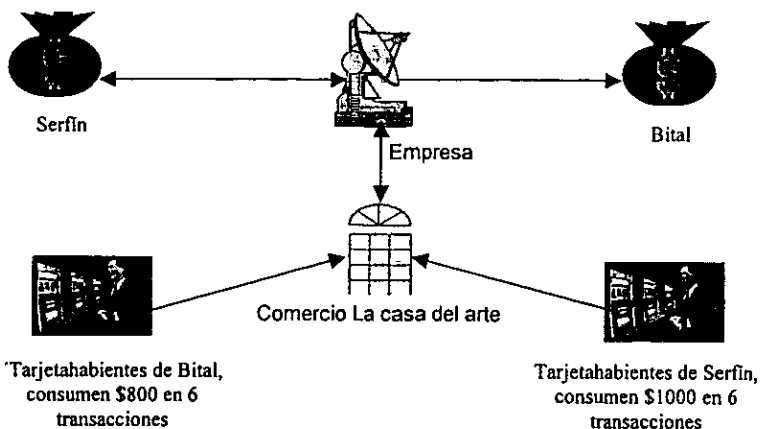
## **Papel del sistema de información estratégica en la empresa**

La empresa debe presentar reportes a los bancos sobre las transacciones realizadas con tarjetas de crédito o débito en los comercios afiliados a los bancos del consorcio, esto con el fin de tener un resumen estadístico del número de transacciones que se realizaron y el monto que se tuvo con las mismas, ya que la empresa debe de realizar una liquidación o compensación entre los bancos.

Por ejemplo, existe un comercio llamado "La casa del arte" el cual se encuentra afiliado al banco Serfin. La casa del arte acepta diferentes tarjetas de crédito y débito.



La siguiente figura muestra de forma general la tarea realizada por la empresa.



La empresa con relación a los bancos y comercios

Como puede observarse, en el comercio se pueden generar transacciones de tarjetahabientes afiliados a diferentes bancos. La empresa debe ir almacenando todas las transacciones realizadas en el comercio vía POS durante un día hábil. Por cada transacción realizada y dependiendo del monto con que esta se haya generado el comercio debe pagar cierta comisión al banco al que se encuentra asociado, esto sin importar de que tipo de tarjetahabiente provenga la transacción.

Si se realiza una devolución, no existe comisión para el banco, sin embargo el banco debe pagar a la empresa cierta comisión por el uso del switch de transacciones (por ejemplo, imaginemos que el negocio genera 10 transacciones, de las cuales todas son devoluciones, el banco debe pagar por cada transacción que se realice sin importar que haya sido o no devolución, sin embargo si se trata de una devolución el comercio no realiza ningún pago).

Posteriormente se debe realizar una liquidación y compensación entre los bancos y comercios. Esto se realiza a través de la empresa, es decir, la empresa paga al comercio lo que se consumió, en este caso \$1,800 menos las comisiones



correspondientes generadas de las transacciones, Bital deberá pagar \$800 ya que fue la cantidad que se generó de las transacciones, y Serfin deberá pagar \$1000 por el mismo concepto, esta compensación la realizan a través de la empresa. Serfin recibirá cierta comisión generada de los movimientos de los tarjetahabientes y pagar cierta comisión a la empresa por sus servicios del switch de transacciones, compensación y liquidación entre banco y comercio.

Diariamente se generan miles de transacciones, las cuales deben separarse por banco con sus diferentes comercios asociados para poder presentar un reporte de la cantidad monetaria y número de transacciones que se generan en los bancos, esto con el fin de que el comercio tenga un reporte de lo que ha pagado y el comportamiento de sus tarjetahabientes para con los diferentes comercios.

Anteriormente estos reportes eran presentados a los bancos en papel y a un nivel muy general (sólo la facturación y número de transacciones que se tenían en un comercio), no se podía entregar un detalle, la fuente de la que se tomaban los datos era de origen distinto, incompleta y con formato poco amigable.

Para este tipo de información se requería de gran cantidad de horas hombre, ya que se tenía que sumarizar cada transacción realizada vía POS en un comercio afiliado a un banco del consorcio (se habla de aproximadamente 450,000 transacciones por banco), por lo cual sólo se entregaba un estimado.

No podía entregarse a los bancos información acerca de sus comercios con mayor facturación, o aquellos que no facturaban, lo cual generaba hacia el banco poca visión del mercado.

Es entonces cuando la empresa ve la necesidad de presentar mayor información a sus clientes. Realiza entrevistas con los mismos para saber las necesidades de información que estos tienen, entre estas necesidades se tiene:





- La información se pedía que fuera confiable, ya que cuando se realizaba un estimado este podía obtenerse de fuentes de información diferentes.
- En los reportes que se solicitaban, se deseaba fuera más ágil su generación, ya que si se pedía un reporte de la facturación de un mes, este debería solicitarse con anticipación de 20 días, cuando el cliente realmente tenía urgencia de verificar sus cifras.
- De parte del cliente poder ver la transacción de sus comercios agrupados en diferentes categorías, tales como población, la cadena de tiendas a la que correspondían y por rubros de restaurantes, líneas aéreas, hoteles, tiendas, y otros.
- Muchas veces era de gran ayuda para los bancos que la información de sus comercios se presentara de forma mas detallada. Por ejemplo, que la facturación de Sanborns se presentara de cada uno de los Sanborns que existen en la república, no en nivel general.
- Para los bancos es importante saber que comercios no se encuentran activos, o cuales son aquellos que tienen perdidas.
- Por ejemplo, aquellos comercios que generaban gran cantidad de devoluciones, ya que para el banco esto podía implicar perdidas.
- Además el banco solicitaba ver cuantas transacciones se habían realizado por medio de sus tarjetahabientes, esto con el fin de ver su comportamiento de estos en sus comercios.

Como resultado de esto, la empresa se ve en la necesidad de crear su sistema de información estratégica (SIE), el cual se alimenta de un Data Warehouse<sup>7</sup>. La información se explota a través de reportes y gráficas, presentando la facturación, transacciones, comercios activos y rentabilidad de los comercios asociados a los bancos del consorcio.

---

<sup>7</sup> Repositorio de datos, el cual contiene información necesaria para la toma de decisiones.



Este DWH sólo contenía información de adquirente, de ahí que se desarrolle del SIE el módulo adquirente, teniendo planeado poblar aún más este DWH con información de cajeros automáticos e información de mercado, para presentar una visión similar a la del módulo adquirente.

En este módulo adquirente se explota la información del DWH, el cual contiene los movimientos que se realizan con tarjetas ya sea de crédito o débito en comercios afiliados a bancos del consorcio y que pasan por el switch de transacciones de la empresa.



## Reflexión de la información general acerca de la organización

Es muy importante señalar los puntos en los que se tuvo participación en el desarrollo del proyecto, ya que de esta se adquirió cierta experiencia para crecer profesionalmente.

Inicialmente para el desarrollo se contaba con los siguientes recursos:

1 Líder de proyecto

1 Analista

2 Programadores

El dar una propuesta a la problemática que tenía la empresa se tuvo poca participación, ya que al inicio del proyecto el papel que se desempeñó fue como analista/programador.

Aunque cabe señalar que para tener participación en el proyecto se tuvo que estudiar el rol de la empresa, pues esto ayuda mucho a tener una mejor visión en los problemas que ésta presente y con esto poder dar sugerencias en la solución, además de dar algunas propuestas para nuevos proyectos.

Conforme se fue avanzando en el proyecto la experiencia ayudó a que se fungiera como líder de proyecto, para lo cual se pide mayor noción de la empresa, así como conocimientos mas firmes en sistemas, puesto que en ese momento se requerirá tomar decisiones sobre el rumbo que tomará el proyecto, así mismo, se requiere tener una buena relación con personal, esto debido a que se tiene a cargo a ciertos recursos.



# Capítulo III.

## Análisis de requerimientos de la empresa

**El problema científico crucial para la investigación de sistemas es éste: cómo separar un sistema viable específico para su estudio del resto del universo sin perpetrar una división aniquilante**

**Stafford Beer**



De acuerdo con los requerimientos de la empresa, se establece la necesidad de construir un DWH, el cual contuviese en primera instancia información acerca de las transacciones realizadas en los comercios asociados a bancos del consorcio. Dichas transacciones son todas aquellas que han pasado por el switch de transacciones de la empresa y que han sido compensadas<sup>8</sup>.

Toda la información asociada al DWH se requería que fuese explotada para realizar reportes, de tal manera que el usuario pudiera hacer sus propias consultas. Para los reportes se solicitó que se mostrará información de:

- Facturación (importe que se obtiene cuando el tarjetahabiente realiza la compra de un bien o servicio, ejerciendo su crédito o sus fondos invertidos, así como las disposiciones de efectivo que haga en su línea de crédito).
- Transacciones (Operaciones o movimientos derivados de la utilización de una tarjeta bancaria, que inciden de manera directa o indirecta en el saldo de crédito).
- Comercios activos (Número de comercios afiliados a un banco del consorcio que registran cuando menos una transacción (venta o devolución) al mes).
- Rentabilidad (Cualidad de una inversión o negocio de proporcionar un rendimiento atractivo).
- Ticket promedio (Valor promedio de las transacciones realizadas).
- Análisis libre (Uso de todas aquellas variables susceptibles a análisis, el usuario podrá hacer uso de toda la información existente y manipularla como considere más conveniente para analizarla y poder hacer una mejor toma de decisiones).

Se pidió que los reportes contuvieran el siguiente formato:

---

<sup>8</sup> Proceso que realizan los adquirentes para saldar cuentas con los emisores y los negocios afiliados derivado de los pagarés emitidos por el uso de una tarjeta bancaria



- Para análisis libre se requería que la información se presentara ordenada por las principales agrupaciones (hoteles, restaurantes, líneas aéreas, tiendas y otros). Este tipo de reporte debería tener la opción de presentar información que el usuario requiriera para su análisis, se tendría la facilidad de elegir información que se considerara importante para la toma de decisiones, de esta forma el usuario no se vería limitado en los reportes prediseñados que se tuvieran. La información que se puede presentar se indicará más adelante.
- Para facturación, transacciones, comercios activos, rentabilidad y ticket promedio, se solicitó lo siguiente:

Al activar cualquiera de estas opciones, se debería mostrar la información requerida agrupada por hoteles, restaurantes, líneas aéreas, tiendas y otros, además se pidió se tuviera opción de presentar los datos en forma gráfica.

Se pedía que la información pudiera ser analizada más a detalle, por ejemplo, que la información agrupada por restaurantes se pudiera visualizar por cadenas (ejemplo Vips, Sanborns, Chano y Chon, etc.) y la información agrupada en cierta cadena se pudiera presentar a detalle de comercio, es decir el nombre del comercio.

Con objeto de facilitar más el análisis se pidió que la información se presentara ordenada de mayor a menor monto en facturación, transacciones, comercios activos, rentabilidad y ticket promedio.

Así mismo, en caso querer visualizar información de facturación y número de transacciones, ésta debería mostrarse agrupada por ventas y devoluciones.

Para visualizar información de rentabilidad, esta mostraría el importe de la comisión (Valor que paga el comercio al banco por el uso de sus servicios) y costo de intercambio (Valor que paga el banco a la empresa por el uso de sus servicios).



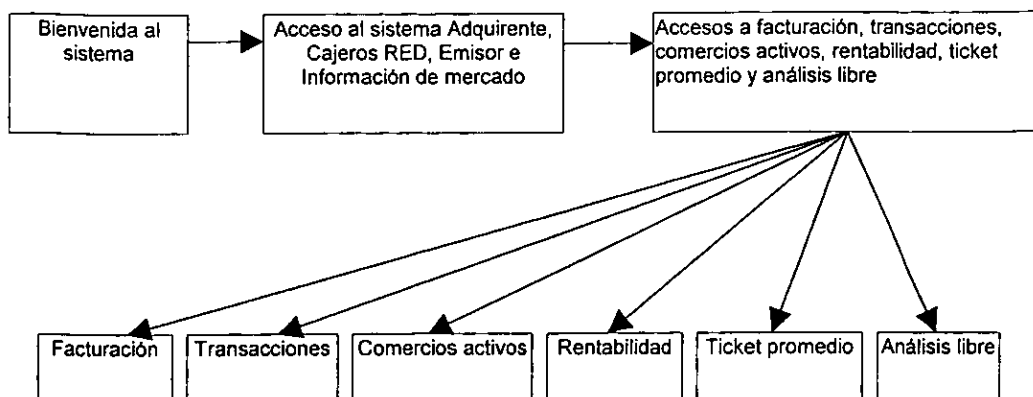
- También se pedía que el usuario tuviera la facilidad de aplicar ciertas condiciones a su información, por ejemplo poder visualizar la forma en que las transacciones son recibidas, es decir, si una transacción fue manual o electrónica (se consideran electrónicas aquellas que son vía POS), o visualizar sólo información en la cual se presenten todas aquellas transacciones que fueron rechazadas o aceptadas, etc.
- De igual forma, se pedía la información mostrada pudiera enviarse a archivos con formato excel.
- Además, para tener una mejor visualización en el negocio de adquirente (banco que asocia a un comercio) se pedía que se mostrara una tendencia sobre la facturación, las transacciones y la rentabilidad que se presentaba en las diferentes agrupaciones.

Para el DWH, las necesidades de la empresa implicaban poblarlo con información de:

- Cajeros RED, presentando todos los tipos de rechazo que se generan en los cajeros, las aclaraciones realizadas, transacciones y monto de las mismas, etcétera.
- Emisor, presentando la facturación que tienen las diferentes tarjetas dependiendo del banco al que correspondan, y el monto que se tiene de tarjetas extranjeras en México.
- Información de mercado, es decir, el comportamiento de los bancos del consorcio en relación con la competencia.



El mapa organizacional de las funciones solicitadas es como sigue:



Organigrama de funciones solicitadas

Se requería presentar al usuario las siguientes interfaces:

Bienvenida al usuario a la aplicación, incluyendo seguridad de acceso, ya que los bancos del consorcio (usuarios) no podrían ver información de otro banco. Una vez teniendo acceso, se presentaría el contenido de la aplicación.

Se deberían mostrar los diferentes módulos con los que cuenta el sistema, es decir, los módulos de Adquirente, Emisor, Cajeros Red e Información de mercado.

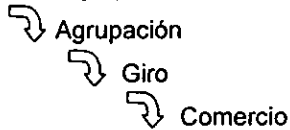
En las opciones que se solicitaron para el módulo adquirente se encontraban facturación, transacciones, comercios activos, rentabilidad, ticket promedio y la parte de análisis libre. Estas opciones deberían presentar un reporte histórico con información de trece meses. Estos trece meses serían de la fecha actual a trece meses de histórico, esto con el fin de que el usuario pudiera visualizar la información de un año y con ello el comportamiento que se tuvo en su negocio, lo cual le podría dar un mejor panorama para la toma de decisiones.





De acuerdo a la forma en que el usuario desea visualizar la información se considera tener dentro de las dimensiones (formas como se pide que se visualice la información) tres grupos, estos son:

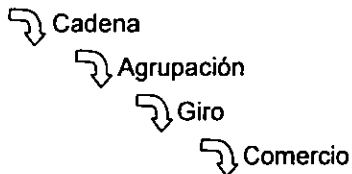
**Primer grupo de dimensiones**



- **Segundo grupo de dimensiones**



- **Tercer grupo de dimensiones**



Donde:

**Agrupación**

Agrupación de los giros, específicamente hoteles, restaurantes, líneas aéreas, tiendas y otros.



## **Giro**

Segmentación de los comercios de acuerdo a su actividad.

## **Comercio**

Nombre del comercio.

## **Población**

Estado en donde se encuentra el comercio.

## **Cadena**

Agrupación de comercios según su tenedor, p.e. El globo agrupando a cada una de sus tiendas, Sanborns agrupando al restaurante, a la farmacia, etc. De cada una de sus tiendas.



## Reflexión del análisis de requerimientos

Para el análisis de requerimientos, el líder inicial del proyecto estudió la situación de la empresa en equipo con el departamento de sistemas de la misma y un consultor, juntos acordaron se construyera un DWH, dada la gran cantidad y poca organización de información que esta tiene.

Como analista de sistemas se recopiló información de datos que se tenían y el tipo de información a la que se deseaba tener un acceso rápido y fácil de consultar, para lo cual se llegó a un acuerdo con el líder de proyecto de separar la información de acuerdo a los temas con mayor interés de tal forma que cumplieran con los requisitos de información sin que la presentación de la misma se prestara a confusión.

El papel del departamento de sistemas era el facilitar información, esta dependería de la que se acordará en juntas, de acuerdo a la forma en que se poblaría el DWH. Para el módulo de adquirente, como ya se mencionó, se requería que las transacciones realizadas en los comercios, fueran depositadas en un archivo texto. La información para llenar dicho archivo es tomada de lo que arroja el Switch de transacciones.

El papel del líder y analista (en el que se tuvo participación al inicio del proyecto) fue el encargarse de la parte de desarrollo, el estudio y agrupación de la información, el diseño e implementación del sistema de información.

Los requerimientos de usuarios fueron tomados a través de juntas, en las cuales participaron el departamento de negocios estratégicos, ya que este es quien mejor conoce la información que puede ser estratégica para los bancos (clientes), el departamento de sistemas, puesto que es quien conoce la información que puede ser presentada o bien que puede ser accesible, el instituto SAS, dado que el



desarrollo se había planeado hacerlo bajo esta herramienta, por lo cual daría propuestas de la forma en que podría ser presentada la información y la forma en que podría hacerse uso a su herramienta, de tal forma que pudiera presentar soluciones en el proyecto (la persona representante de SAS fungió como líder inicial del proyecto), un consultor, ya que la empresa no se dedicaría al desarrollo del sistema se apoyó de una consultoría, la cual debería de estar presente para dar una propuesta al planteamiento del problema y posteriormente se dedicaría a realizar el desarrollo de acuerdo a los requerimientos presentados.



# Capítulo IV.

## OLAP (On Line Analytical Processing)

Excepto por accidente, rara vez sería posible para un individuo maximizar más de un valor al mismo tiempo. También es raro que alguien tenga sólo una meta y la quiera maximizar sin tomar en cuenta su efecto sobre los demás objetivos, a pesar del concepto contrario del hombre económico. Esto significa que los tomadores de decisiones muy pocas veces buscan maximizar un solo objetivo.

Alfred Kuhn



## Historia

La evolución de la informática de los últimos años ha traído consigo una acumulación masiva de datos. Esto proviene en su mayor parte de la aplicación de la informática en las actividades transaccionales de la empresa tales como contabilidad, facturación, gestión de datos, etcétera. Toda esta cantidad de datos carece de valor para el análisis y toma de decisiones sin una organización previa, por lo tanto, hace falta una forma de convertir estos datos en información útil, tal que aporte una nueva perspectiva de los datos. De este modo nace la tecnología OLAP.

La descripción del término OLAP (On Line Analytical Processing, Procesamiento Analítico en Línea) fue acuñado por Boyce Codd, quien estableció las reglas que constituyen la base de la generación actual en tecnología de bases de datos relacionales.

Para satisfacer esta necesidad algunas empresas de software se unieron para buscar una arquitectura que pudiera unir las bases de datos relacionales con la tecnología multidimensional<sup>9</sup>. Esto llevó a nuevos requerimientos en las herramientas de apoyo a la toma de decisiones, ya que ahora estas herramientas deberían incluir aspectos de base de datos relacionales y multidimensionales.

La primera herramienta para la creación de aplicaciones OLAP que se desarrolló fue el lenguaje de programación APL (Analytical Processing Language) que, mediante el uso de datos multidimensionales, otorgaba a sus aplicaciones una funcionalidad similar a la de los productos OLAP de hoy en día. Fue creado por Ken Iverson hacia los años 70 e implantado por IBM. El APL tuvo

---

<sup>9</sup> Datos almacenados de manera lógica en arreglos.



poca aceptación debido a la dificultad en su uso incluso para los programadores experimentados.

A principio de los años 80 Comshare's System W. Construyó un producto que utilizaba datos multidimensionales A este producto le siguieron otros como Commander Prism, Essbase, Metaphor, etcétera. Desafortunadamente el standard de computadoras personales de aquella época no era capaz de soportar este tipo de aplicaciones, ya que éstas sólo podían usarse en máquinas de gran capacidad.

Por esta misma época nace el concepto EIS (Executive Information System, Sistemas de Información Ejecutiva), siendo Pilot Commander Center uno de los primeros programas que adoptan este nombre. También aparecen las hojas de cálculo con apariencia multidimensional.

Los sistemas OLAP son EIS utilizados para proporcionar información estratégica fiable sobre los indicadores del funcionamiento de una organización. OLAP no es una herramienta de usuario, sino un entorno de desarrollo de aplicaciones que están destinadas a solventar un abanico de necesidades de información y poner los medios para obtener una información depurada, confiable y ajustada a nuestras necesidades, ya que un SIE es un sistema de información que permite a los ejecutivos acceso rápido y efectivo a información compartida utilizando interfaces gráficas fáciles de consultar y con ello una rápida interpretación de la información. También proporcionan mecanismos para consolidación de información. Permite al usuario crear sus propias consultas, es decir, los reportes no están fijos, lo cual permite una fácil modificación en el diseño del sistema para adecuaciones particulares.

La tecnología OLAP permite una gran rapidez en el análisis de los datos. Esta rapidez se traduce en información depurada, ya que si no utilizáramos una herramienta OLAP nos limitaríamos a observar los datos de uno en uno, es decir ver cada uno de los registros existentes en todas las bases de datos, buscando



cierta información que nos interesase de algún tema específico e iríamos obteniendo estadísticas o gráficas de los datos que nosotros creyésemos más interesantes o ilustrativos.

Con una herramienta OLAP podemos en el mismo tiempo comparar multitud de datos y gráficas además de extraer relaciones que a simple vista no se nos hubiese ocurrido.

Algunas áreas en las que se hace uso de la tecnología OLAP son: análisis de ventas, rentabilidad y calidad, además de la realización de informes financieros e informes de gestión de datos.

La gran mayoría de los datos que se usan en aplicaciones OLAP son originarias de otros sistemas operacionales y aplicaciones. En casi todas las aplicaciones OLAP los datos son capturados directamente por la aplicación. Cuando los datos proceden de otras aplicaciones es necesario duplicarlos con el fin de almacenarlos por separado de los datos originales, ya que se requiere que estos sean utilizados de manera activa e independiente. Los datos muy comúnmente hacen referencia al Data Warehouse (DWH), ya que los datos que se depositan en éstos son utilizados para consulta y análisis (Se ahonda acerca de esto en el tema de los DWH). Estos datos son de gran confianza, ya que se realiza en ellos una depuración de la información, a diferencia de los sistemas de procesamiento de transacciones en línea (OLPT por sus siglas en inglés), en donde los datos se reúnen y almacenan para operación y control.

## **Características de OLAP**

Como ya se mencionó anteriormente, una de las características distintivas de OLAP estriba en la duplicación de los datos. Las razones por las cuales es





necesario que se dupliquen los datos para formar el DWH del que se va a alimentar la aplicación OLAP son:

### **Ejecución**

Las aplicaciones OLAP son con frecuencia de un gran tamaño y se suelen usar para realizar análisis iterativos inciertos, es decir, realizar análisis de cierta información al momento de ejecución, sin que este se encuentre predefinido, este tipo de reportes si no existiera la técnica OLAP, carecerían de rapidez en la consulta, ya que como las aplicaciones son por lo regular de gran cantidad de datos, el presentar la información tardaría mucho.

Esto requiere que se pueda acceder a los datos de manera muy rápida, lo cual obliga a que se guarden por separado y a disponer de una estructura de datos optimizada que pueda ser accedida sin perjudicar la respuesta operativa del sistema.

### **Múltiples fuentes de datos**

Muchas aplicaciones OLAP requieren datos que son originados desde distintos y diversos sistemas, tanto procedentes de fuentes externas como de algunas aplicaciones del escritorio. El proceso para unir y combinar estos datos procedentes de distintas aplicaciones o sistemas puede ser extremadamente complejo, ya que estas aplicaciones o sistemas suelen usar métodos de codificación diferentes y además pueden tener periodicidades distintas. Es por ello que los datos se ponen por separado de los sistemas, ya que se requiere realizar cambio en formatos y ponerlos en mismas periodicidades, además que como los sistemas operacionales cambian de un momento a otro, se requiere que en una consulta los datos sean los mismos. Los métodos de codificación no se realizan en



el momento, estos se realizan antes de la consulta para que de esta forma esta sea rápida y eficiente.

### **Filtrado de datos**

En la gran mayoría de los sistemas transaccionales nos encontramos con mucha frecuencia gran cantidad de datos erróneos que necesitan ser filtrados antes, para poder realizar un buen análisis que nos permita generar informes adecuados.

En caso de que no se filtraran los datos podrían aparecer ciertas distorsiones a la hora de realizar un análisis, pudiendo llevar a la toma de una decisión equivocada. Por ejemplo, si no se filtran nuestros datos y en el sistema operacional del que se está alimentando el OLAP no se llenan ciertos campos, o tal vez se registran doble vez, nos llevará a tener datos erróneos, con lo cual nuestra información carecería de validez.

### **Ajuste y modificación de datos**

Algunas de las razones por las cuales se deben de ajustar los datos son:

- Las distintas estructuras de la compañía no siempre son iguales, ya que existen diferencias al trabajar en direcciones departamentales.
- Se pueden realizar análisis que no parten de datos operativos como pueden ser los que se obtienen de características demográficas, publicidad televisiva, etcétera, con lo cual se tendría que realizar ciertas modificaciones, ya que probablemente se introduce información que no es válida para el tipo de campo en el que se está almacenando.
- También puede ser que existan sucursales situadas en otros países y que operen con contabilidades distintas, por lo tanto podría ser seguro que los datos deban ser modificados antes de realizar un análisis. Por ejemplo, hablemos de un producto el cual el precio en un país es en dólares y en



México es en pesos, se debe realizar un cambio o modificación en los datos, quedando el producto con un formato de precio igual.

### Actualización y consistencia de datos

Si los datos que se van a utilizar en una aplicación OLAP provienen de diferentes fuentes, es probable que no se puedan actualizar todos al mismo tiempo, ya que podrían venir de diferentes sistemas operativos y estos seguro tendrían diferentes estados de actualización. Por ejemplo, al final del mes puede ser que los datos de las ventas estén al día, ya que la aplicación así lo trabaja, sin embargo, los datos de fabricación se encuentran actualizados hasta la segunda o tercer semana del mes, ya que provienen de otro sistema operacional.

También, como las empresas utilizan varios sistemas operacionales seguramente se tienen datos que tienen diferentes formatos, puesto que en un sistema a un campo se le puede llamar de una manera diferente que en otro, o incluso el tipo de dato puede tener diferente formato aunque los datos se refieran a lo mismo. Por ejemplo, imaginemos se tiene un campo que hace referencia a montos, estos pueden tener un formato de pesos en un sistema, mientras que en otro no contener ningún formato.

Por ejemplo, pueden existir 3 sistemas (s1, s1 y s3) que manejen diferentes nomenclaturas con respecto al sexo, tal como lo muestra la siguiente figura:

S 1		S 2		S 3	
Clave	Descripción	Clave	Descripción	Clave	Descripción
M	Masculino	H	Hombre	2	Masculino
F	Femenino	M	Mujer	1	Femenino

Nomenclaturas diferentes en tres sistemas diferentes



Esto hace necesario estandarizar dichas nomenclaturas para lograr la consistencia de los datos, tal como se muestra a continuación:

Sexo OLAP	
Clave	Descripción
M	Masculino
F	Femenino

Nomenclatura estándar para OLAP

### Historia de los datos

Muchas aplicaciones OLAP toman al tiempo como una dimensión, ya que es muy importante para la toma de decisiones tener pronósticos o comparativos de tiempo. Para obtener datos de años anteriores es importante aplicar gran esfuerzo, ya que se requiere de migrar los datos de aplicaciones antiguas y ajustarlos para que puedan ser utilizados en la base de datos OLAP.

### Distintas perspectivas o vistas

Los datos operacionales tienen que ser necesariamente muy detallados, pero muchas de las actividades de toma de decisiones requieren de una visión más amplia, es decir, de más alto nivel, no tan estructurada. Es por ello que se deben combinar varios almacenes de datos, ajustar la información según el nivel de resumen o nivel de visión que se desee ver.

Se tienen varias formas de almacenar los datos para que posteriormente sean extraídos y utilizados en el análisis por herramientas OLAP, algunos de estos tipos de almacenamiento pueden ser Bases de Datos Relacionales (ROLAP (Relational On Line Analytical Processing)) o Bases de Datos Multidimensionales (MOLAP (Multidimensional On Line Analytical Processing)). Estos se describen a continuación.



## **Base de datos relacionales OLAP (ROLAP)**

El nombre ROLAP se aplica a los productos OLAP que toman una base de datos relacional como soporte para la extracción de la información.

En un ROLAP, los datos se encuentran físicamente almacenados en un RDBMS (Base de datos relacional por sus siglas en inglés),

No se tiene la habilidad de realizar cálculos multidimensionales usando relaciones simples. En la mayoría de los casos se permite un número limitado de cálculos en SQL y los resultados se usan como entrada de un motor multidimensional, que es el que realiza la mayoría del trabajo.

El modelo entidad-relación por lo regular es usado cuando los datos para la construcción del DWH son tomados de varios sistemas operacionales. En este contexto, este modelo se encuentra basado en la teoría de conjuntos y SQL, en alta normalización (mínima o sin redundancia de datos) y la optimización para la actualización de datos. El uso de éste modelo requiere de complejas combinaciones en las consultas.

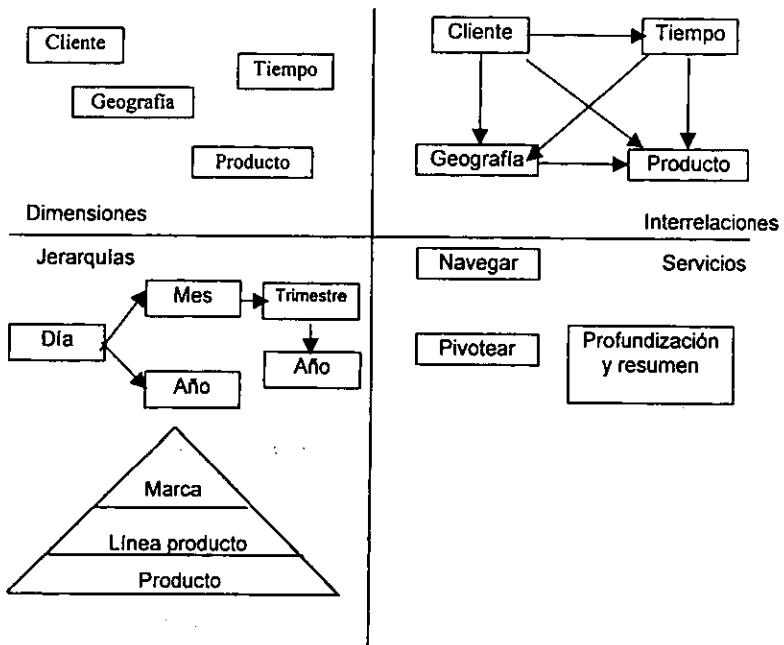
## **Base de datos multidimensionales OLAP (MOLAP)**

El nombre MOLAP se aplica a los productos OLAP que toman una base de datos multidimensional como soporte para la extracción de la información.

Estas bases de datos multidimensionales suelen adquirir datos de otras fuentes como base de datos relacionales, herramientas de escritorio u hojas de cálculo, también algunos datos son introducidos por usuarios finales.



Para entender mejor como se almacenan los datos en una base de datos multidimensional, debemos saber como es el análisis multidimensional, la siguiente figura muestra este análisis.



Análisis multidimensional

La figura anterior muestra como en el análisis multidimensional los datos se presentan mediante dimensiones, por ejemplo producto, territorio (geografía) y cliente. Por lo regular las dimensiones se relacionan en jerarquías como por ejemplo para la dimensión territorio, se puede tener una jerarquía de ciudad, estado, región, país, continente. El tiempo también es una dimensión estándar con su propia jerarquía como día, mes, trimestre y año.



Los resultados que se tienen en una consulta (servicios que nos da la multidimensionalidad) se pueden pivotear o girar para cambiar los ejes y con ello la perspectiva, es decir, de las dimensiones que se dispongan se pueden interrelacionar para obtener mayor información o bien, obtener información desde otro tipo de vista. Los usuarios pueden navegar por las dimensiones profundizando u obteniendo resúmenes a lo largo de los elementos de una dimensión (esto a través de las jerarquías), o penetrar a través de las dimensiones para ver otras perspectivas.

Ahora bien, una base de datos multidimensional (MDDB) es un diseño que almacena dimensiones agregadas en n-formas de combinaciones.

Se forma una matriz o cubo de dimensiones, donde cada dimensión está compuesta por un número de niveles. Por ejemplo, imaginemos que tenemos los valores para visualizar información de facturación neta en una cadena de tiendas, y esta información la podemos visualizar por lugar y tiempo.

La facturación por lugar la podemos ver por: ciudad, estado y población. A esta dimensión la llamaremos "dimensión de lugar".

La facturación por tiempo la podemos ver por: año, semestre y mes. A esta dimensión la llamaremos "dimensión de tiempo".

La combinación por nivel de dimensión forma una celda. Cada celda contiene un vector de valores agregados para cada cruce en particular. En este caso, una celda podría estar compuesta por país, otra por país y ciudad, otra por país y estado, otra por país y población, etcétera.



## Consejos al crear un MDDB

Se debe tener cuidado con los siguientes aspectos al generar un MDDB:

Una dimensión no se puede agregar a un MDDB ya creado, esta deberá ser agregada al momento de construirla.

La omisión de una dimensión en la estructura del MDDB forzará al usuario a remitirse al detalle de los datos, lo que claramente afectará el desempeño en la consulta que se realice.

El siguiente ejemplo explicará la forma en que se representa una base de datos multidimensional:

Imaginemos una fábrica que produce pantalones. La empresa tiene un análisis multidimensional en el cual se tienen las dimensiones de productos, geografía y clientes.

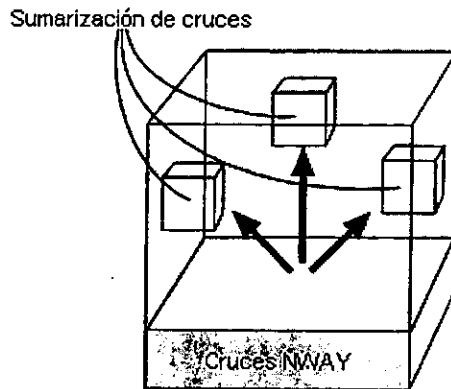
Cada dimensión tiene un determinado número de niveles, por ejemplo:

Productos tienen los niveles de marca, línea de producto

Cada nivel de agregación puede ser almacenado en cualquier celda.

El siguiente dibujo muestra la forma en la que los datos pueden ser almacenados en una base de datos multidimensional.





Almacenamiento de una base de datos multidimensional

La mayoría de aplicaciones OLAP se encuentran orientadas para uso interactivo de tal manera que el usuario recibe una rápida respuesta a sus interrogantes. Esto suele cumplirse si la respuesta se reduce a reunir información de la base de datos, darle formato y presentarla al usuario, pero resultará muy lento si se requiere de hacer grandes cálculos para tener una respuesta. Este costeo de tiempo no es por los cálculos aritméticos que se requieren, sino se debe a la recuperación de datos que afectan a los elementos que se van a calcular.

Para poder conseguir una respuesta rápida, todas las aplicaciones multidimensionales deben precalcular algunos datos que se vayan a utilizar en el análisis. En las bases de datos multidimensionales el almacenamiento de datos precalculado suele ser automático y transparente, mientras que ROLAP normalmente se usan tablas resumen.



El modelo de tablas de sumarización almacenan temas categorizados por dimensiones tales como tiempos y región. Un tema puede tener un gran número de tablas sumarizadas asociadas a él.

Frecuentemente estas tablas de sumarización siguen jerarquías naturales que pueden realizar drill down (bajar de nivel) y up (subir de nivel) en la información sumarizada.

Una desventaja de este modelo es que si se omite la sumarización de una tabla, se forzará al usuario a tener referencia al detalle de los datos, es decir, si el usuario realizara una consulta en el DWH y la tabla a la que hace referencia a dicha consulta no existe se tendrá que realizar una sumarización de todos los datos al momento, lo cual claramente afectará el rendimiento en las consultas.



## Reflexión de OLAP (On Line Analytical Processing )

Dado que el requerimiento de la empresa fue el generar información estratégica se tenía que hacer uso de la herramienta OLAP, pues se pretende que los usuarios tengan los medios necesarios para tener acceso rápido a la información.

Como ya se mencionó, se hizo uso de archivos TXT a través de los cuales se realizaron las extracciones de información apropiada, esta fue inicialmente guardada en una tabla temporal ya que posteriormente se realizarían modificaciones que se requirieran, esto debido a las uniones con otras fuentes de datos.

En la etapa de transformación y manipulación de datos se fungió como líder de proyecto, teniendo a cargo como recursos a un programador y un analista, cabe mencionar que estos recursos se tuvieron en la etapa de desarrollo del módulo adquirente, posteriormente en el módulo de cajeros automáticos, en el cual se anexó un programador y para el módulo de competencia de mercado sólo fue necesario un programador como recurso.

Ya que se deseaba tener información mensual de la facturaciones realizadas en comercios (Adquirente), fue necesario pedir las transacciones realizadas mensualmente, dado que la cantidad de información que se tiene es muy grande.

Como líder se tomó la decisión de que los archivos se entregaran todos los días, con lo cual el usuario tendría información oportuna de doce meses de histórico así como parte del mes en curso, no se tendría que esperar a que finalizara este para poder contar con información estratégica que le ayudase a conocer la situación de sus comercios.



Por parte de sistemas no se tendrían que generar procesos muy pesados de forma que afectarán el rendimiento de otras aplicaciones.

Como fuentes de datos se tenían los comercios que fungían como adquirente, giros que estos tienen, movimientos realizados en comercios (vía POST), poblaciones en las que estos se encuentran, cadenas a las que corresponden, así como las interedes que tienen (en caso de que se presente).

Para poder adquirir esta información se tuvo que estar en las áreas de negocios estratégicos y sistemas ya que al realizar la unión de los datos se encontraba cierta distorsión por lo que fue necesario hacer un filtrado de los mismos teniendo como apoyo a las dos áreas mencionadas para verificar la información.

Para el filtrado de los datos se fue realizando uniones por partes verificando el resultado de cada una de tal forma que cuando ya no existiera incoherencia en la información, entonces se realizaba la unión de todas las fuentes de datos.

Dado que se tenía acceso a diferentes fuentes, los datos tuvieron que homogeneizarse ya que por ejemplo, el archivo que entregaba el área de sistemas con las transacciones realizadas hacía mención de interredes a través de un rango de claves, por ejemplo, los valores 1 y 3 correspondían a la interred liverpool, mientras que para el área negocios estratégicos sólo se hacía la distinción por el nombre de la interred.

Para las poblaciones, un área entregaba a los comercios y sus poblaciones, mientras que en el archivo de facturaciones se encontraban las claves de los estados, por lo que se tuvo que homogeneizar la fuente de comercios, adaptándola a cierto estado.



Al archivo de facturaciones se tenía que hacer cierta operación sobre las comisiones que se cobran a los comercios de acuerdo al monto de la transacción, ya que la información correspondiente a este campo se tomaba algunas veces como importe y otras veces como porcentaje de comisión.

Una vez que los datos quedaron homogeneizados, se generó un proceso automático el cual realiza la carga de la información ciertos días de la semana, para esta carga se tomarían el archivo texto (txt) con las facturaciones realizadas por los comercios, así como catálogos necesarios, algunos de estos eran actualizados semanalmente, para lo cual existía otro proceso automático de carga, puesto que por ejemplo, podrían darse de alta nuevos comercios en la empresa, en los cuales empezaría a existir facturación, por lo que sería necesario saber información del comercio como nombre, intered, cadena a la que pertenece, etc.

Para el tipo de almacenamiento el líder inicial del proyecto ya había tomado la decisión de crear bases de datos multidimensionales, debido a que esta es una de las herramientas con las que cuenta el producto SAS, y como se mencionó, el líder inicial es representante del instituto SAS.

De acuerdo a la experiencia a través del proyecto, se sabe que algunas de las razones para el uso de bases de datos multidimensionales fueron la gran cantidad de datos y cálculo requerido para la presentación de información así como jerarquización de la misma a través de explosión de datos.

Debido a que en esta etapa del proyecto ya se fingía como líder, para tomar una buena decisión en el diseño del cubo y la forma en que estarían las tablas base para la construcción del mismo fue necesario el estudio de bases de datos multidimensionales.



Se tuvo cuidado en no omitir presumarizaciones, ya que esto costaría rendimiento a la aplicación. Para ello se tomaron las diferentes jerarquías que se tenían en los requerimientos y conforme a estos se tomaron las posibles combinaciones que pudiesen hacerse, de tal forma que en las consultas estuvieran presentes las presumarizaciones requeridas.



# Capítulo V.

## Data Warehouse (DWH)

**En el orden de las ideas radica la armoniosidad de los hechos y la capacidad para comprender la sabiduría de la vida.**

**O. Delgadillo (Azúl)**



## Introducción

Hoy en día las empresas compiten cada vez más, por lo que se requiere información valiosa con la cual se puedan crear estrategias empresariales y tomar decisiones que lleven a situar al negocio por delante de los competidores. Una de las herramientas de las que hoy en día se dispone se denomina técnicas OLAP (Procesos analíticos en línea por sus siglas en inglés), con lo que se puede hacer un análisis multidimensional de los datos, el cual consiste en organizar la información de acuerdo a los parámetros que los altos directivos consideren convenientes para dar un sentido a la información y realizar un análisis profundo.

Cuando se va a utilizar OLAP es necesario que los datos activos sean almacenados por separado de la aplicación OLAP. Estos datos por lo regular pueden referirse a un DWH o Datamart.

## Definiciones de Data Warehouse

Algunas definiciones de lo que es un Data Warehouse son:

Un Data Warehouse es una estrategia y un método para separar de manera lógica y física los datos operativos de los datos para la toma de decisiones.

Apuntes del curso DWH del ITESM.

Un Data warehouse (DWH), es un repositorio separado de datos, en el cual estos están almacenados en una forma adecuada para los sistemas de inteligencia de negocios y soporte a decisiones, de tal manera que los sistemas no interfieran con el desempeño de los sistemas operacionales.

SAS Institute





Un Data warehouse es un conjunto de datos orientados a un tema, los datos están integrados y son variables en el tiempo, es decir, se tiene un histórico, no volátiles. Estos datos son empleados como apoyo en la toma de decisiones administrativas.

W. H. Inmon

Las definiciones interpretan lo mismo: Un data warehouse es un repositorio de datos, y estos contienen información necesaria para la toma de decisiones.

La diferencia que existe entre un sistema Operacional y un DWH es que los sistemas operacionales estándares tales como ventas y distribución, reservaciones de vuelos, administración de materiales y muchos otros, no difieren demasiado de una compañía a otra, debido a que no requieren de un conocimiento profundo del enfoque del negocio específico de cada compañía; por lo tanto es relativamente fácil entregarlos a un subcontratista.

Los DWH son diferentes en todas las compañías y requieren de un gran volumen de información de las mismas. Para ello se requiere de la limpieza de los datos, lo cual no es fácil, ya que debe existir un análisis muy profundo, el cual tiene que asegurar que no existen anomalías ni problemas con la información, por lo tanto, todos los datos pasarán por un proceso de homogeneización y validación.

El departamento de sistemas, al estar más cerca del negocio de lo que pueda estar un subcontratista, es el proveedor natural de un DWH, es por ello que el DWH ofrece al departamento de sistemas una oportunidad para ayudar a la gente de las áreas de negocios a obtener sus objetivos de largo plazo, ya que el DWH es la infraestructura que contiene el fundamento para una correcta toma de decisiones por parte de los trabajadores del conocimiento (knowledge workers) en los departamentos de usuarios finales, y en último término por parte de la dirección.



El DWH puede servir a un grupo de personas y departamentos que se relacionan a través de la información, de tal manera que puede cumplir de manera satisfactoria intereses muy diferentes entre sí derivados del uso que se le dará a los datos.

Es por ello que un DWH puede organizar los datos sobre un tema específico como puede ser clientes, productos, fuerza de ventas entre otros y compartir esta información con varios departamentos.

Para la implementación del DWH es necesario la participación del usuario final, ya que es él quien conoce del negocio y la información que requiere para toma de decisiones.

Las actividades necesarias que debe tener todo DWH son: administración, organización y explotación y que serán explicadas más adelante.

## **Características técnicas de un DWH**

- Un DWH está organizado por temas. Un tema regularmente requiere de datos que deben ser tomados de diferentes sistemas operacionales, por ejemplo: Productos, clientes, garantías, vuelos, cargos, empleados, cursos, etcétera. Refleja las necesidades del negocio, ya que los componentes del DWH serán puntos específicos que reflejen directamente el tema. Algunos de estos puntos específicos podrían ser regiones, productos o clientes. Estos puntos deben ser definidos para satisfacer las necesidades del negocio, no para conocer los requerimientos de los sistemas operacionales.
- La información que se encuentra en un DWH es una colección de datos tomados en intervalos de tiempo regulares, ya que los datos históricos son importantes en un sistema de soporte a decisiones.



- Los DWH son usados para el análisis, reportes y consultas de grandes volúmenes de datos.
- Un DWH no es volátil, es decir, si dos personas desean consultar la misma información, deberán obtener la misma respuesta. También significa que los datos no cambiarán durante el análisis, ya que estos son de solo lectura.
- El DWH es Integro. Es decir, toda la información que proviene de los sistemas operacionales deberá ser consistente en los diferentes puntos específicos del DWH. La diferencia que exista entre los datos deberá ser primero validada antes de ser pasada al DWH, de tal manera que la información sea segura y no existan incoherencias en la misma.
- Un DWH está optimizado para tener un buen desempeño en las consultas, para ello, por lo regular se requiere de presumarización y preindexación de datos, además de habilidad para separar y combinar los datos en varias formas posibles.
- Cuenta con herramientas para análisis, ya que un DWH estará completo si existen herramientas para la toma de decisiones que reflejen y exploten la estructura del DWH. Por lo regular, para esta explotación se utilizan los sistemas de información ejecutivos.

Pueden existir varias razones para crear un DWH, pero se debe tener bien claro las necesidades de información que el negocio requiere para construir el DWH y los criterios para evaluar el éxito del proyecto.

## **Pasos para la creación de un DWH**

Los pasos que se siguen para la creación de un Data Warehouse son los siguientes:



### **Justificación**

Se construye el equipo del proyecto, formado por el departamento de sistemas y usuarios finales. El equipo comienza evaluando el posible impacto del DWH y el porqué se justifica el inicio de dicho proyecto.

### **Recolección de requerimientos**

El equipo empieza a recolectar requerimientos de información de los diferentes departamentos de los usuarios finales, examina la arquitectura y la estrategia del área encargada de tecnología de información (IT).

### **Diseño/prototipo**

Se definen reglas de transformación de los datos recolectados. Se realizan prototipos de extracción de los datos integrados y transformados, esto en forma de programas muestra. Se hacen pruebas de carga de datos y se realizan consultas de muestra y otros tipos de reportes que utilizará el usuario final.

### **Implementación**

Se escriben todos los programas y aplicaciones necesarios y se pone el DWH a disposición del usuario final.

### **Revisión**

Se evalúa el proyecto, tomando puntos buenos e implicaciones del mismo para el negocio de la compañía. Esto se realiza en forma inmediata y posteriormente se vuelve a evaluar una vez transcurrido cierto tiempo.



Cada una de las fases anteriores es muy importante, ya que cada una provee una entrada lógica a la fase siguiente. Seguramente se cree que la fase de implementación es una de las más importantes, sin embargo, esto es erróneo. Los DWH no se desarrollan mediante una implementación de explotación, se van desarrollando mediante una serie de proyectos iterativos, en los que cada uno va agregando algo.

A continuación los puntos mencionados serán mostrados más a detalle utilizando la metodología de Rapid warehouse del instituto SAS. Las fases se repiten a través de varios ciclos como un proyecto de Desarrollo Rápido de Aplicaciones (RAD por sus siglas en inglés).

## **1. Justificación**

Un DWH requiere de la experiencia de varias personas de la compañía, estas deberán representar las diferentes funciones de los departamentos afectados en el proyecto del DWH.

El objetivo fundamental de esta etapa es que la dirección superior y todos los demás grupos participantes estén de acuerdo en la creación del DWH y lo aprueben.

La primera tarea que se debe realizar es evaluar en los departamentos de usuarios finales y en Sistemas el posible impacto que tendrá el DWH en el negocio y demostrar el porqué se justifica la creación del mismo.

El equipo del proyecto deberá de proveer material que pueda justificar la creación del DWH. Se deberán visitar organizaciones (preferentemente de la misma



industria) que hayan implementado exitosamente un DWH. También se deberán revisar artículos de prensa y encuestas de analistas que hablen del DWH.

El material que debe entregar el equipo del proyecto para la justificación del DWH deberá incluir:

- Una explicación de lo que es un DWH.
- Lo que no es un DWH, es decir, dejar bien claro que no son aplicaciones operacionales, con el fin de evitar ideas equivocadas de lo que pueda esperarse del mismo.
- Los beneficios que obtendría la compañía en actividades críticas tanto a corto como a largo plazo.
- La manera en que trabajarán el departamento de sistemas y los departamentos de usuarios finales.
- La manera en que cada persona pueda contribuir en el triunfo del proyecto del DWH. Por ejemplo, los departamentos de usuarios finales pueden proporcionar información clave acerca de las necesidades de explotación del DWH y el departamento de sistemas puede proporcionar experiencia en administración y organización del mismo.

La comunicación es un elemento clave en todas las fases, pero principalmente en esta, ya que se deberá encontrar una justificación clara para la creación y aceptación del DWH.

Como ya se mencionó, se deberá presentar una justificación y además una propuesta para la creación del DWH. Una vez que se hayan entregado estos documentos a la dirección, se deberá tomar la decisión en cuanto a proseguir o no. En caso de ser afirmativa, cuál sería el alcance inicial del proyecto (duración, recursos, departamentos de usuario final involucrados, etcétera). El alcance de duración inicial deberá reducirse a datos, es decir, que información deberá



incluirse en el DWH al inicio ya que si para apoyar la explotación inicial que se propone se requiere del 90 por ciento de los datos corporativos, entonces estamos hablando de un proyecto muy ambicioso para tratarse en el primer ciclo de implementación.

Los productos que deberán generarse al final de esta etapa son:

- Documento de propuesta: se especifica lo que se pretende hacer, el porque de la creación de un DWH, los beneficios que este tiene.
- Presentación de justificación: Presentar los beneficios que traerá para la empresa.
- Alcance inicial del proyecto: Limitar lo que se quiere abarcar en cuanto a información.
- Definición del equipo de implementación (departamentos de usuarios finales): Es decir, aquellas personas que estarán involucradas en el proyecto, ya que de ellas depende el tema sobre el cual se poblará el DWH.

## **2. Recolección de requerimientos**

Al estar en esta etapa, ya se ha obtenido un alto grado de compromiso y participación.

El propósito de esta fase consiste en definir los temas objetivo de que se poblará el DWH y definir la forma en que se encontrará el Data Warehouse lógica y físicamente.

Se deberá de convocar a una junta de planeación con el fin de trazar el proceso que se seguirá y asegurarse de que los departamentos de usuarios finales participantes tengan una idea clara de lo que será requerido de ellos en esta fase y de que todos concuerden en alcanzar los objetivos trazados. Así mismo, esta



La junta dará oportunidad de plantear y obtener respuesta de todas las preguntas que hayan quedado en el aire en la fase anterior.

Se incluirá un proceso de entrevistas para definir de forma más precisa los requerimientos de los usuarios. Se tendrán entrevistas con gerencias superiores, gerentes de los departamentos de usuarios finales del Data Warehouse y el departamento de sistemas. Se debe tener muy presente que el proceso de entrevistas bien ejecutado es muy importante para el éxito de la implementación de un DWH.

Como parte de los procedimientos de entrevistas, el equipo del proyecto deberá revisar los reportes existentes, análisis de datos y sistemas de soporte a decisiones. Esto contribuirá a una mejor comprensión de los puntos a ser atacados y resueltos en el diseño de los datos del DWH, por ejemplo información inadecuada, inconsistencias, duplicación de esfuerzos, etcétera.

Las entrevistas llenarán un documento de Definición de los requerimientos. En el cual, se proveerá un resumen detallado de las necesidades de los usuarios finales, las necesidades de arquitectura/estrategia del departamento de sistemas, las definiciones de los temas, los modelos lógico y físico del Data Warehouse, los modelos de transformación y un alcance del trabajo de implementación.

Un elemento del documento de definición de requerimientos incluye el Modelo lógico, el cual define las entidades de los datos que se requieren para resolver la(s) necesidad(es) de negocio que está(n) siendo atacada(s).

Para definir el modelo lógico se sugiere lo siguiente:

- Obtener la funcionalidad deseada del DWH.





- Confirmar las áreas de temas de los datos. Por ejemplo cliente, producto, etcétera.
- Identificar la variante del tiempo y la granularidad de tema(s), es decir con cuánto tiempo se tendrá la información y qué tan a detalle se tendrá.
- Descomponer los temas de los datos en entidades de datos abarcando hechos y dimensiones.
- Dar el análisis de los atributos detallados de los datos a la fase de modelación física.

Una vez que se tiene el modelo lógico, este deberá ser plasmado en un modelo físico de datos, el cual definirá la arquitectura real de almacenamiento del DWH y las tablas que contendrá. La definición de las tablas incluye definición de columnas en cada tabla, atributos de las columnas fuentes de datos operacionales de las columnas, definición de índices, especificación de llaves (primarias y secundarias), plataforma en que se encontrarán las tablas y espacios estimados para cada tabla.

El diseño físico deberá tener en cuenta la forma en que se esperan usar los datos y la manera de organizarlos dependiendo del uso más frecuente que se pretenda hacer de los mismos. Este es un punto que deberá ser pensado muy cuidadosamente, ya que debe tener en cuenta la importancia de extraer información del DWH a partir de consultas y reporte de datos. También deberá tenerse en cuenta cuáles *data marts* serán definidos. Los *data marts*, también llamados Warehouses de aplicación, son extractos o resúmenes específicos de ciertos requerimientos. Quizá a primera vista esto parecería contradecir el objetivo de hacer el Data warehouse, sin embargo, lo que se busca con definir los *data marts* es simplemente el ser prácticos en lo que respecta al desempeño, optimización de almacenamiento y requerimientos locales.

El modelo físico de datos para la construcción del DWH puede estar basado en varios tipos de diseño, estos pueden ser: Modelo entidad-relación, esquemas de



estrella o copos de nieve, almacenamiento multidimensional persistente y tablas de sumarización. Un DWH puede utilizar más de uno de estos. Para cualquiera de las partes del DWH es aconsejable usar la forma más simple que sea capaz de soportar los requerimientos.

Finalmente, una vez que se tiene el modelo físico se iniciará con el modelo de transformación. Aquí se definirá cómo se transferirán los datos operacionales al almacenamiento del DWH. Como los datos provenientes de los sistemas operacionales por lo regular se encuentran en una forma inconsistente para el soporte a las decisiones, será necesario un proceso de transformación de los datos. Esta transformación consiste en tres pasos: conversión, enriquecimiento y sumarización.

#### **a) Conversión**

Está dirigida a resolver las inconsistencias de los datos, esto hablando en términos de valores y formatos entre diferentes sistemas operacionales. Por ejemplo, imaginemos que un sistema operacional, en su base de datos utiliza los valores de "F" para determinar el sexo femenino y "M" para el masculino, mientras que en otro sistema operacional, se utiliza 1 para determinar femenino y 2 para masculino. Estos valores deberán hacerse consistentes antes de transferirlos a un ambiente de DWH.

#### **b) Enriquecimiento**

Consiste en combinar diferentes atributos de diferentes sistemas operacionales con el fin de crear nuevas entidades para el análisis del DWH. Por ejemplo, imaginemos que en un sistema operacional se captura el precio de venta de un producto X y en otro sistema operacional se captura la utilidad bruta del producto. Ahora imaginemos que en el DWH quisiéramos analizar el margen bruto, una



transformación adecuada sería dividir la utilidad bruta del producto por el precio de venta del mismo.

### c) Sumarización

En el ambiente del DWH la sumarización es crítica, desde la perspectiva de proporcionar al analista una visión histórica en lugar de una visión de un registro a la vez, como es el caso de una base de datos de un sistema operacional. Los datos pueden ser presumarizados antes de ser cargados en el warehouse (metadata), lo cual ayudaría a reducir los volúmenes de datos, o pueden ser sumarizados para un determinado propósito desde dentro del warehouse. Los resúmenes consisten tanto de sumarizaciones numéricas como de agrupamientos o conteos.

Al finalizar esta fase se obtendrá un documento de definición de requerimientos.

## 3. Diseño/prototipo

La fase anterior provee la base para la fase de diseño/prototipo, cuyo propósito es determinar la estructura y contenido del DWH y definir cómo se transformarán los datos operacionales para probarlo.

Esta fase debe comenzar por el desarrollo conjunto de aplicaciones, donde el objetivo de esta sesión deberá ser, con la participación de todo el equipo del proyecto, el revisar el documento de definición de requerimientos. Además se deben realizar las siguientes actividades:

- Validar el proceso de entrevistas.
- Reconciliar las necesidades de los usuarios de negocios con las restricciones de la organización.



- Evitar en esta fase y en la fase de implementación, los errores costosos e iteraciones innecesarias.
- Producir el modelo lógico y físico del DWH.
- Editar el modelo de transformación.

El equipo del proyecto iniciará el proceso iterativo para diseñar e implementar el DWH, realizando un prototipo y éste se utilizará para probar los modelos lógico, físico y de transformación. El prototipo se mostrará a los usuarios del DWH, dándoles oportunidad de que retroalimenten al equipo del proyecto.

Se generará la lógica del programa para probar diferentes fases de la implementación del DWH. Las fases incluyen la adquisición de datos transformación de los datos, transferencia y carga de los datos y su explotación.

El prototipo será la base a partir de la cual el equipo del proyecto puede continuar hasta completar la construcción del DWH. Este prototipo puede consistir únicamente en una prueba del concepto para el modelo de datos y las herramientas de explotación, sin tener en cuenta los aspectos de calendarización /actualización, que deberán ser considerados en la fase de implementación.

Al finalizar esta fase se deberá tener completo como ya se ha mencionado:

- Un modelo lógico del DWH.
- Un modelo físico de los datos que permita a los desarrolladores de la aplicación un progreso rápido en la fase de implementación.
- Un modelo de transformación de datos para mapearlos de los sistemas operacionales en el warehouse.
- Entendimiento de la forma en que los requerimientos de los usuarios encajan en los modelos físico y lógico del DWH.
- Un prototipo del DWH.



## **4. Implementación**

En esta fase se debe construir un ambiente del DWH enfocado a tres aspectos: Administración, Organización y explotación.

### **a) Administración**

Se refiere a cómo se relacionarán la lógica de extracción, transformación y carga, con el fin de lograr una entrega oportuna del DWH con datos de calidad.

### **b) Organización**

Se refiere a la administración y mantenimiento del almacenamiento de datos tanto lógico como físico.

### **c) Explotación**

Se refiere a la interfaz de usuario del ambiente del DWH que involucra la selección, toma y entrega de las herramientas de acceso al mismo y la capacitación del usuario final.

### **d) Acceso**

Dentro de la fase de acceso, se contemplan los siguientes aspectos:

#### **i) Acceso a Datos**

Para poblar el DWH, el primer paso será extraer los datos del ambiente operacional. Estos datos podrían encontrarse en cualquier formato, siendo el trabajo en esta parte la forma en que se extraerían.



## ii) Integración y transformación de los datos

Los datos que se encuentran almacenados de una forma adecuada para los sistemas operacionales deben ser integrados y transformados para que se encuentren orientados a temas, sean consistentes y no volátiles. Es por ello que debe existir cierta lógica de manejo de datos operacionales, agregándoles valor antes de colocarlos en el DWH, para que de esta forma no se aplique dicha lógica al realizar la explotación ya que así el usuario tendría que aplicar dicha lógica al realizar sus consultas, lo cual no sería muy apropiado, ya que se le debe facilitar lo mas posible la presentación de su información. De esta manera se tomará una aplicación en forma de estructura de datos que se encontrará orientada a negocios y no a aplicaciones.

## iii) Calendarización

En todo DWH existen tareas que deben llevarse a cabo en intervalos regulares, tales como:

- Actualizaciones periódicas del DWH.
- Distribución de tablas desde un DWH central a Data Warehouses departamentales o data marts, toda vez que el DWH central es actualizado.
- Generación periódica de reportes tabulares o gráfico, los cuales deben de presentar el nivel (status) del DWH, es decir lo que se ha actualizado, si la actualización ha sido correcta, resúmenes de información que se tiene, etcétera.

En un DWH exitoso estas tareas deben de automatizarse lo mas posible, de tal manera que se reduzca la carga de trabajo del administrador del DWH. Esto se



debe realizar a través del calendarizador del DWH, el cual debe automatizar la tarea de someter los trabajos (Jobs) mediante una interfaz de usuario independiente de la plataforma. El calendarizador consiste en:

- Una interfaz de usuario para que sea el mismo quien calendarice los trabajos.
- Tablas de control del calendarizador que contengan información acerca de los trabajos calendarizados, por ejemplo que trabajo debe de correr en que plataforma y a qué hora.
- Un servidor calendarizador que “se despierte” a intervalos predefinidas de tiempo para ver las tablas de control y someter los trabajos a las horas ahí establecidas (El calendarizador podría ser, hablando de Unix un daemon, que se le indique la realización de ciertas tareas a determinado tiempo, o un cron<sup>10</sup>).

#### **iv) Carga del Data Warehouse**

Una vez que se ha terminado con la transformación de los datos, se puede cargar el DWH. En el diseño del DWH se debe de especificar cuan a menudo debe de actualizarse a partir de la carga inicial, a esto se le denomina renovación, esta debe ser calendarizada. Posterior a la carga inicial existen dos posibles enfoques con relación a la renovación del DWH, estos son:

**Renovación completa.** En la mayoría de los casos debido a los volúmenes crecientes de información, no es posible una carga continua del DWH desde el inicio.

---

<sup>10</sup>Herramienta que permite ejecutar procesos calendarizados , esto a través de comandos en Unix, esta herramienta permite que automáticamente los procesos sean ejecutados. (definición completa ver anexo).



El segundo es la carga de los datos que han cambiado, es decir, cargar al DWH de los datos operacionales únicamente aquellos que han cambiado.

Aquí el punto más importante es la probable complejidad de correr o examinar los sistemas operacionales fuente de manera que los cambios puedan ser capturados con exactitud. Las posibles opciones incluyen:

- Examinar los datos internos seleccionando los registros con criterios que se hayan modificado desde la última renovación.
- Usar archivos para escribir todos los cambios que se han hecho.
- Generar listados de estatus (Logs) sobre los llamados a las bases de datos, estos se registran en archivos por separado.

#### **v) Explotación del Data Warehouse**

Es de vital importancia disponer de información estructurada y organizada para poder satisfacer las necesidades de negocios y el DWH es un medio para lograrlo. Mientras no se disponga de herramientas que permitan al usuario final examinar, analizar y producir reportes que apoyen una mejor toma de decisiones el DWH estará incompleto.

Dependiendo de los requerimientos de los usuarios finales, las herramientas de explotación del DWH pueden ir desde herramientas simples y listas para usar consultas y hacer reportes, pasando por herramientas de análisis multidimensionales hasta aplicaciones de Sistemas de Información Ejecutivos (SIE) avanzadas diseñadas para satisfacer objetivos específicos de la compañía.





## 5. Revisión

Como se había mencionado, los DWH son construidos en forma incremental y por lo tanto es importante aprender de los éxitos y errores y documentar estos últimos para la siguiente vuelta de la implementación incremental. Después de haber completado la implementación, se deberá hacer una revisión y los resultados deberán ser entregados al administrador del DWH y a los departamentos de usuario final. Se debe hacer una segunda revisión entre seis y nueve meses después de haber completado la implementación. Es opcional realizar una tercera revisión después de 18 a 24 meses de la implementación, esto con el fin de observar las implicaciones a largo plazo del DWH.

A medida que se realicen estas revisiones, se puede asegurar la introducción de nuevos temas.

El proyecto de un DWH generalmente se encuentra diseñado para satisfacer un amplio rango de requerimientos de información de uno o varios departamentos. Este mismo DWH sirve a diferentes propósitos de entrega de información, algunos de los cuales pueden no quedar claros hasta que el proyecto se encuentre encaminado. Tanto los departamentos de usuario final como los gerentes se encuentran involucrados en el proyecto, y los beneficios del mismo afectan directa o indirectamente a muchos o tal vez a la mayoría del personal de la organización.



## Reflexión del Data Warehouse

En la etapa inicial del DWH no se estuvo presente, sin embargo, cuando se empezó a tener participación se tenía que saber como estaban las etapas anteriores.

En la parte de justificación se encontraron los siguientes participantes:

- Departamento de sistemas de la empresa
- Instituto SAS como apoyo en el uso de su herramienta y además quien inicialmente fungió como líder del proyecto
- Un consultor
- Departamento de negocios estratégicos de la empresa

Los participantes mencionados se dedicaron a presentar una justificación del porque la creación de un DWH, inicialmente hicieron encuestas para ver si la empresa estaba en condiciones de la creación del DWH.

En la parte de recolección de requerimientos se tuvo participación como analista / programador al pasar el modelo lógico a un modelo físico. Ya que se tenían definidas las entidades en datos y dimensiones, es decir ya se tenían los campos que debería contener algunas entidades, además de las dimensiones que se solicitaron de acuerdo a los requerimientos. A partir de esto se tuvo que analizar la forma en que quedarían las tablas, pues se tenían ciertas especificaciones para algunos campos, por ejemplo, se sabía que existía un campo correspondiente a la cadena de algún comercio, pero no se sabía que diferentes cadenas podrían existir, y el formato en que se encontraban en las tablas.

En el modelo de transformación, como se señaló en la reflexión anterior, se realizaron transformaciones en los datos para que no existieran inconsistencias en



los mismos, para esta transformación se tuvo apoyo de las áreas involucradas en la creación del DWH.

Para el diseño del DWH se realizó un prototipo con datos de un mes de las transacciones generadas a través del Switch de transacciones de la empresa. A partir de este punto se fueron depurando los datos que tomaría al final el DWH.

El que se generará un prototipo no significa que se toman los datos tal como se entregan, primero se debió hacer una carga de estos, de los cual se encontraron inconsistencias, se fue trabajando en ellas hasta conseguir que la información estuviera correcta. Una vez que se tenían los datos listos, se realizó la prueba de almacenarlos en bases de datos multidimensionales, esto para evaluar si era posible este tipo de almacenamiento de acuerdo a las peticiones realizadas por los usuarios.

En esta parte del de diseño del DWH se aplicó OLAP, pues de aquí que se toman datos de aplicaciones operacionales y se realizan transformaciones a los mismos.

Para la explotación del DWH se consideró el uso de sistemas de información ejecutivos, estos presentarían la información que se pidió en los requerimientos, de tal forma que pudieran navegar a través de ella , realizando explotaciones de la misma dependiendo en el nivel en que se encontrara. Para esto fue necesario la creación de cubos (Base de datos multidimensional).

El diseño final que se entregó referente a los sistemas de información ejecutivos, puede verse en el anexo del presente trabajo.



# Capítulo VI.

## Herramienta SAS

**Ningún hombre, sabiéndolo, escoge la peor  
de dos posibilidades**

**Pitágoras**

**La libertad empieza cuando la ignorancia  
acaba**

**José Alfredo López**



## Historia

SAS Institute fue fundada por el Dr. James H. Goodnight, su presidente, con el fin de desarrollar un producto de análisis estadístico totalmente acabado. Desde entonces SAS Institute ha introducido productos de software integrados que responden a cualquier necesidad en la toma de decisiones en los negocios y sus soluciones están presentes en el sector Industrial, organismos gubernamentales y en el ámbito académico.

El producto original de SAS Institute fue desarrollado para analizar información proveniente de investigaciones agrícolas en un Mainframe IBM en la Universidad de Carolina del Norte.

A medida que cambió la industria, también lo hizo el sistema SAS. En la década de los 80's el sistema SAS fue totalmente reescrito en lenguaje C, a fin de incorporarle la MultiVendor Architecture (MVA), lo cual permite soportar nuevas plataformas al poco tiempo de liberarlas. Debido a la MVA, el 90% del código SAS es transportable a cualquier plataforma.

En 1993, la estrategia MVA permitió a SAS Institute liberar en forma casi simultánea nuevas versiones para diferentes plataformas: MVS, CMS, VSE, ApenVMS, para VAX y AXP, OS/2 2.0, Windows 3.1, Windows NT, AIX, HP-UX, RISC/ULTRIX, Solaris y ConvexOS.

El sistema SAS es un sistema de software modular integrado e independiente del hardware, diseñado para satisfacer todas las necesidades de procesamiento y entrega de información que las organizaciones requieran por grande y compleja que esa sea. Se distingue por su habilidad para:

- La transformación de los datos de la empresa en información significativa.



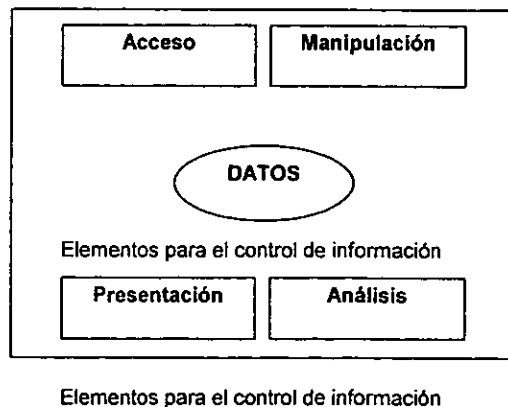
- Poner esa información a disposición de quienes la necesitan.
- Procesar los datos en forma consistente a través de un amplio rango de ambientes de hardware, aprovechando las ventajas de cada uno de ellos.

SAS permite un completo control sobre los datos, desde el acceso, manejo y análisis hasta su presentación. El sistema SAS incluye herramientas para desarrollo de aplicaciones orientado a objetos, capacidades avanzadas de procesamiento cliente/servidor, nueva tecnología de visualización de datos y acceso ilimitado a los mismos.

Está diseñado para el control de datos, de tal forma que puede tomarlos en cualquier formato y en cualquier lugar, permitiendo a los usuarios manipularlos de acuerdo a sus necesidades.

Para poder tener un control total de los datos y generar información útil y manejable, se requiere de cuatro tareas importantes ya mencionadas:

Acceso, manipulación, análisis y presentación.





### **1.- Acceso**

Permite acceder a los datos requeridos por las aplicaciones, sin importar como o donde se almacenen o el formato que tengan.

### **2.- Manipulación**

Permite manipular los datos para actualizarlos, combinarlos, editarlos, o realizar subconjunto de los mismos antes de realizar el análisis.

### **3.- Análisis**

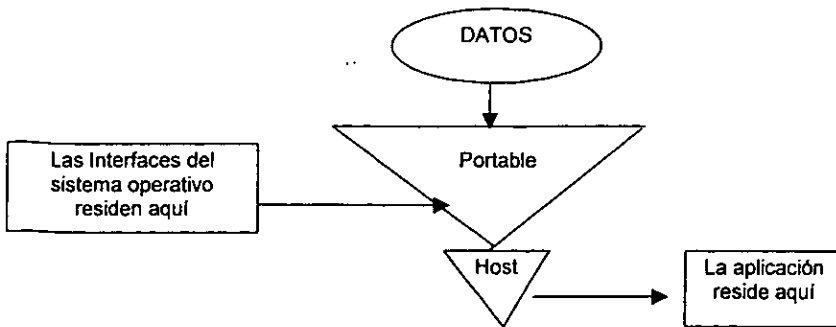
SAS cuenta con una gran variedad de herramientas para análisis de datos, estas van desde simples estadísticas hasta análisis especializados.

### **4.- Presentación**

Esta se puede realizar por medio de tablas o elaboración de diferentes tipos de gráficas.

SAS permite que las aplicaciones sean portables, esto se debe a su estructura MultiVendor Architecture (MVA). Mucha de la funcionalidad del sistema SAS esta contenida en un componente portable, mientras los componentes del host proveen todas las interfaces necesarias para el sistema operativo y hardware. De esta forma, sin importar el hardware o sistema operativo, las aplicaciones SAS funcionarán igual y producirán los mismos resultados tanto en una computadora personal como en un mainframe.

El siguiente diagrama muestra en forma general como se encuentra la estructura Multivendor en SAS.



Estructura Multivendedor de SAS

El componente host del sistema SAS es escrito por separado de cada ambiente y permite al sistema usar la tecnología que proveen los diferentes sistemas operativos.

El beneficio que se obtiene de tener aplicaciones portables es que puede ejecutar la misma aplicación en los ambientes que se requiera. Es decir, el ambiente de computo esta dado por el hardware que se utiliza y el sistema operativo que corre en el. Se pueden desarrollar aplicaciones con SAS en un ambiente y ejecutarlo en este, o en cualquier otro sin necesidad de rehacer la aplicación.

El lenguaje SAS y sus procedimientos hace posible la realización de una gran variedad de aplicaciones para propósitos de análisis, acceso, presentación y manipulación de datos.

Como SAS provee control total de los datos, por lo tanto se pueden desarrollar aplicaciones de cualquier necesidad para explotación y análisis de los mismos.

Entre las aplicaciones que se pueden desarrollar se encuentran las siguientes:

- Soporte en la toma de decisiones





- Sistemas de información ejecutivos
- Reportes y análisis financieros
- Administración de proyectos
- Análisis de laboratorio de datos

Hoy en día, en una empresa u organización, existen grandes cantidades de datos, estas pueden encontrarse en diferentes lados y diferentes formatos, ya que pueden pertenecer a diferentes aplicaciones o departamentos.

El sistema SAS puede acceder a diferentes sistemas de bases de datos tales como: IBM Corporation's DB2, Digital Equipment Corporation's Rdb/VMS, Oracle corporation's ORACLE.

Los datos que son extraídos de las diferentes bases de datos por SAS, son almacenados en Data sets SAS (tablas propias de SAS).

Una vez que se obtiene el data set SAS, entonces se pueden hacer diferentes procesos para la realización de la aplicación.

Algunos de los productos o módulos que se integran en el sistema SAS son:

### **SAS/Access**

Es una serie de interfaces individuales que proveen enlaces directos y transparentes sobre lectura/escritura y una gran variedad de administración sobre diferentes bases de datos.

### **SAS/AF**

Provee ambiente para desarrollo de aplicaciones orientadas a objetos.



Este módulo puede ser utilizado para crear diversos tipos de interfaces que utilicen menús.

### **SAS/Assist**

Es un ambiente para desarrollo de aplicaciones vía point and click. Este módulo permite desarrollar ciertas tareas con el mínimo de esfuerzo, ya que SAS se encarga de generar el código correspondiente para la tarea deseada.

### **SAS/Calc**

Es una hoja de cálculo con capacidad para la manipulación de información. Se puede hacer uso de ésta hoja de cálculo para manipular datos que se encuentren en formato SAS, de algunas bases de datos o hasta de archivos planos.

### **SAS/Connect**

Provee conexión cliente/servidor en un ambiente SAS que permite al cliente ejecutar el sistema SAS para establecer comunicación con una o más aplicaciones SAS o programas que se encuentren corriendo en ambientes remotos.

### **SAS/EIS**

Son aplicaciones orientadas a objetos, las cuales ya se encuentran desarrolladas para creación y mantenimiento de sistemas de información. Ofrece código libre, lo cual puede demostrar la capacidad que tiene el sistema SAS para acceso, manipulación, análisis y presentación de información.

### **SAS/ETS**

Es una herramienta de análisis para series de tiempo, econometría y predicción, planeación y modelación financiera. Incluye procedimientos para el análisis de



series de tiempo, simulación lineal y no-lineal, amortización, depreciación y generación de reportes financieros.

### **SAS/FSP**

Permite interactuar con los datos, editarlos, actualizarlos, etcétera. Se pueden simular diferentes formas para presentar los datos, crear tablas que muestren los datos para una rápida actualización.

### **SAS/Graph**

Provee información y presentación gráfica a color para producir una serie de gráficas, histogramas y mapas en una gran variedad de colores.

### **SAS/Insight**

Es una herramienta iterativa para explotación y análisis de información, la cual permite a los usuarios examinar, distribuir y explorar los datos a través de histogramas, diagramas de dispersión, diagrama de caja y diagramas en tercera dimensión.

### **SAS/Share**

Permite que múltiples usuarios tengan acceso a datos y actualización a un mismo tiempo.

### **SAS/Stat**

Provee capacidad estadística para análisis regresivo, análisis para datos categóricos, análisis de varianza, análisis multivariado, análisis de cluster, análisis de sobrevivencia y análisis no paramétrico.



### **SAS/Toolkit**

Permite a los usuarios generar sus propios procedimientos, rutinas para acceso interno en el desarrollo de las bases de datos o rutinas especiales.

### **SAS/Multidimensional database**

Permite generar bases de datos multidimensionales para manejo de gran volumen de información y generación de jerarquías en los datos.

Algunos de los módulos mencionados anteriormente se describirán de forma mas detallada, estos son: Bases de datos multidimensionales y HOLAP.

## **Base de datos multidimensional SAS**

El sistema SAS ofrece una base de datos multidimensional (SAS/MDDDB server), la cual es un almacenamiento especial en donde los datos pueden ser tomados desde un DWH o cualquier otras fuentes de datos. Estos datos se almacenan en una especie de matriz, lo cual permite un acceso rápido y fácil debido a las herramientas de acceso multidimensional.

Este modelo de datos para procesos analíticos en línea (OLAP) permiten una mejor toma de decisiones a los usuarios, ya que permite la visualización rápida, con muchas relaciones, ilimitadas vistas de datos sobre datos ya sumariados.

El MDDDB de SAS utiliza una tabla llamada NWAY y diferentes subtablas. Esto se define como sigue:

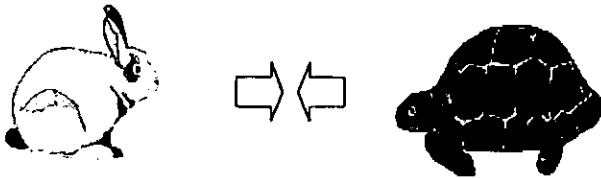
- La tabla NWAY representa la tabla almacenada en el MDDDB que contiene todas las variables (campos) que se han definido. Sólo existe una tabla NWAY en un MDDDB y siempre debe de estar presente.



- Una subtabla es una pequeña colección de datos resumizados que generalmente se derivan de la tabla NWAY. Esta subtabla nace de las diferentes combinaciones de información que se pueden generar, a estas combinaciones se les llama cruces.

Las subtablas son creadas para obtener mayor velocidad en las consultas de información, pero se debe tomar en cuenta que entre mas tablas se generen, mayor es el espacio que se ocupa, por ello al crear un MDDB se debe tomar la decisión de el número de subtablas que se generarán.

La tabla NWAY con varias subtablas o pocas subtablas se podría representar como sigue:



Representación de la tabla NWAY de pocas a varias subtablas

Por lo tanto podremos decir que un MDDB es un simple archivo con una estructura que se encuentra formada de una lista de tablas, donde la tabla más general recibe el nombre de NWAY.

Por ejemplo, imaginemos que tenemos información sobre todas las ventas que se han realizado en una compañía X (por lo regular se realizan 100 ventas por mes), los datos se encuentran divididos por región.



Si se realizara un MDDB, la tabla NWAY sería:

Año	Cuarto de año	Mes	Ciudad	Región	Subregión	Ventas
1994	Q1	1	Mex	Centro	1	34,144
1994	Q1	1	USA	Centro	1	43,567
1994	Q1	1	Mex	Centro	2	64342
.	.	.	.	.	.	.
.	.	.	.	.	.	.
1996	Q4	12	Mex	Sur	5	43,213
1996	Q4	12	USA	Sur	5	83,562

Tabla NWAY

Una subtabla podría ser:

Año	Cuarto de año	Mes	Ventas
1994	Q1	1	234,144
1994	Q1	2	243,567
1994	Q1	3	564,342
.	.	.	.
.	.	.	.
.	.	.	.
1996	Q4	12	543,213

Subtabla

Todo MDDB contiene:

- No limitaciones en el número de subtablas.
- No limitaciones en el número de variables de clase (campos por los que se desea ver información).
- No limitación en el número de variables de análisis (campos por los cuales se analizarán las variables de clase).



- Un máximo de ocho estadísticas.
- Trece estadísticas posibles en tiempo de ejecución.
- Cruces de datos no nulos.
- Capacidad de creación en lotes y en línea.

Un MDDB puede almacenar sólo ocho estadísticas diferentes. Dependiendo de las estadísticas declaradas, se pueden derivar otras trece, la siguiente tabla muestra las estadísticas que se tienen y las que se pueden derivar de estas:

Estadísticas deseadas	Estadísticas Requeridas
N	N
Sum	Sum
Sumwgt	Sumwgt
Uwsum	Uwsum
Nmiss	Nmiss
Uss	Uss
Min	Min
Max	Max
Avg	N, Sum
Range	Min, Max
Pctn	N
Pctsum	Sum
Css	N, Sum, Uss
Var	
Std	
Stderr	
Cv	
T	
Prt	
Lclm	
Uclm	

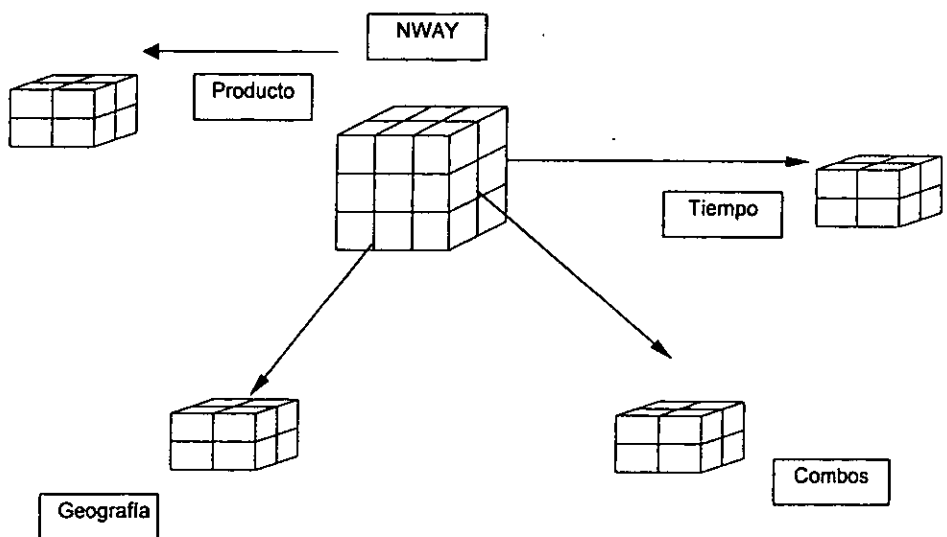
Estadísticas en un MDDB



Almacenando sólo una estadística y cruces (combinaciones) válidas, pueden ayudar a que el tamaño del MDDB sea el mínimo y eficiente.

La creación de un MDDB en el sistema SAS no es difícil, lo difícil es decidir el balance entre la respuesta de una consulta y el espacio de almacenamiento.

El siguiente diagrama muestra un ejemplo de la estructura de un MDDB:



Estructura de un MDDB

En la tabla NWAY se almacenan todas las variables definidas.

Existen  $n$  - número de subtablas que contienen datos resumizados, estos provienen de la tabla NWAY. Estos datos pueden contener varias variables definidas, lo que permitirá que cuando se realice una consulta sobre las variables definidas en una subtabla, los datos ya se encuentren resumizados.





Antes de construir un MDDB, se debe de tener presente lo siguiente:

- Los datos y cómo serán utilizados, ya que conociendo los datos, permitirá que se realice una mejor decisión sobre lo que se debe y no sumarizar además de como categorizar los datos.
- El problema de la empresa, ya que ayudará a decidir prioridades en la construcción del cubo, pues este debe de cumplir con la estrategia del DWH de la empresa.
- Conocer las expectativas del usuario, ya que esto ayudará a decidir la creación de varias subtablas o sólo la tabla NWAY.

Si no se conocieran los datos ni el problema de la empresa, se podría tener un resultado con archivos sumariados muy largos y espacio del archivo muy grande.

### **Procedimiento MDDB**

Existen diferentes formas de construir un MDDB en el sistema SAS, estas son:

- Procedimiento MDDB
- A través del Software EIS de SAS
- A través del Software AF de SAS
- A través del Software Warehouse de SAS

Sólo se mostrará el procedimiento MDDB. Este contiene cuatro sentencias, la sintaxis es la siguiente:

```
Proc mddb <data = nombre del dataset> out = Librería.mddb;  
        <label=descripción para el mddb>;  
    class Variables de categoría  
    var Variables de análisis/estadística (diferentes opciones);
```



**Hierarchy** lista de variables de clase /<name=nombre> <display=YES|NO>

**Run;**

**Proc:** Es usado para especificar la entrada de un archivo que contendrá el detalle de sumarizaciones, este archivo será un MDDB. El data = es usado para especificar el nombre de la tabla base que contiene los datos y el out= el nombre del MDDB que se generará.

**Class:** Especifica las variables de la tabla base con las cuales se realizarán las diferentes combinaciones (a partir de esta se genera la tabla NWAY).

**Var:** Especifica las variables de análisis.

**Hierarchy:** Se especifican las diferentes tablas (subtablas a crear, es decir las diferentes combinaciones de consulta de información que se podrían requerir).

## **HOLAP (Hybrid On line Analytical Processing)**

HOLAP se define como OLAP (Procesos analíticos en línea) Híbrido, esto combina procesos de OLAP relacional y OLAP multidimensional.

El HOLAP es una optimización de las aplicaciones construidas con bases de datos multidimensionales, SAS/EIS y que corren en un ambiente cliente servidor. Incluye la habilidad de ejecutar cálculos y proceso remotos, además permite el acceso a una variedad de estructuras de datos, estos datos incluyen:

- MDDB
- Datasets



- Tablas RDBMS
- Star schemas

Las bases de datos multidimensionales ofrecen un rápido y eficiente acceso a los datos. Este acceso es rápido, ya que el almacenamiento ha sido especialmente diseñado para consultas predefinidas. Esto significa que los índices y el almacenamiento están específicamente diseñados para hacer uso de datos ya resumizados sobre clasificaciones. Por eso mismo son eficientes. Este tipo de base de datos son usadas como de sólo lectura. El manejo de este tipo de bases de datos multidimensionales hacen que sean de un uso variado, ya que para los cambios no requieren de un administrador de base de datos.

Otra ventaja del HOLAP es la funcionalidad que tiene en el servidor. Con HOLAP, el servidor es quien genera todos los subconjuntos de tablas y cálculos y no se generan en el cliente como se hace en las bases de datos multidimensionales. El realizar los cálculos en el cliente puede producir un rendimiento pobre, esto debido al tráfico que puede existir en la red.

Las principales características del HOLAP incluyen:

- Transparencia en la habilidad para realizar procesos en el servidor, esto con el fin de mejorar el rendimiento de la aplicación.
- Habilidad para usar datos distribuidos para incrementar la escalabilidad.
- Cache para dar respuesta inmediata a las pre-sumarizaciones requeridas.
- Log de salida para ayudar a los desarrolladores a detectar cuellos de botella y tener un mejor entendimiento del funcionamiento del HOLAP.
- Mayor flexibilidad en el compartimento de llamadas entre la máquina local y el servidor.
- Mejoramiento en la inicialización de tiempos.



Los beneficios en el uso de HOLAP son:

- Mayor rendimiento.
- Mayor escalabilidad.
- Mayor almacenamiento y distribución en los datos.

El HOLAP es un suplemento de los productos SAS /EIS y SAS MDDB. Fue diseñado para resolver problemas de rendimiento y escalabilidad cuando se hace uso del MDDB y EIS del sistema SAS en un ambiente cliente servidor. Estos problemas son los siguientes:

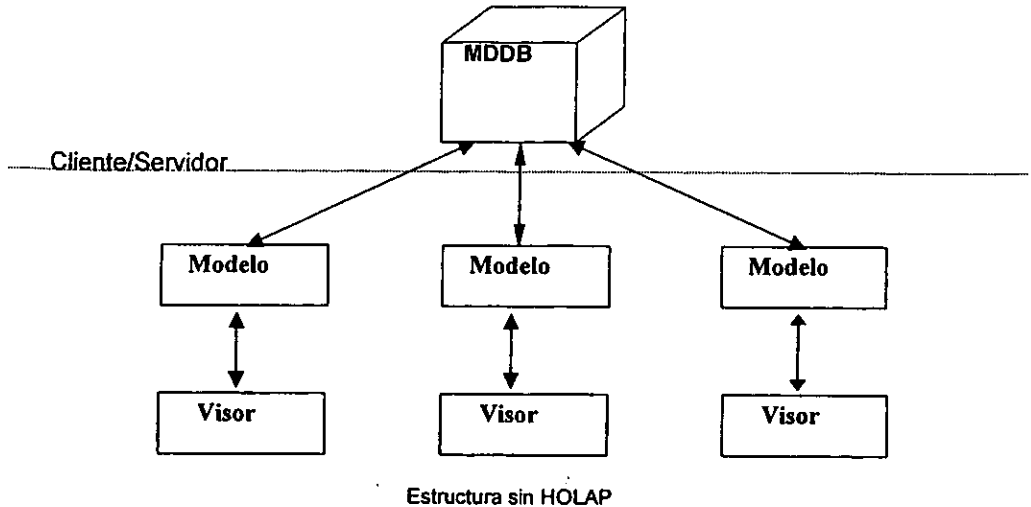
- La inicialización de los objetos multidimensionales toma mucho tiempo.
- Grandes tablas son enviadas del servidor a la memoria de disco del cliente, lo que algunas veces provoca que la memoria no sea suficiente y se provoque la inhibición de la máquina.
- La escalabilidad esta limitada al MDDB que se requiere.
- No es soportada la presumarización de los datos en archivos no MDDB .

## Arquitectura HOLAP

Consideremos un EIS de SAS sin HOLAP, tal como se muestra en la siguiente figura, este tiene un MDDB, el cual se encuentra declarado con servicio de librerías remotas<sup>11</sup>, este MDDB trabaja a través del visor-modelo que se encuentra en el cliente. Es decir, el modelo y visor de los datos que se presentan se encuentran en el cliente y el MDDB, que es donde se encuentran mis datos se encuentran en el servidor.

---

<sup>11</sup> Recordemos que una librería es el alias de la ruta donde se encuentran mis datos o bien, la dirección lógica de los datos. Una librería remota, es aquella que se encuentra declarada en el servidor.



El visor sabe lo que se le desplegará al usuario. Esto puede ser:

- Un reporte multidimensional
- Una gráfica tridimensional
- Un Mapa
- Una gráfica organizacional
- Navegadores de dimensiones
- CSF (Critical Success Factor (Factores crítico de éxito))

Las interfaces del modelo que se encuentran entre el visor y la fuente de datos entre otras cosas sabe:

- El acceso de los datos
- Drill Down (Explotación de datos)
- La navegación
- Rotaciones
- Subconjuntos



Las consultas que se generan en el visor son enviadas al modelo, el cual accede a los datos y baja la información para que el visor la despliegue.

Por ejemplo, consideremos lo que pasa cuando un MDDB se crea en un servidor Unix con el siguiente código:

```
Proc MDDB data = server.master out = server.exmddb;  
  Class year quarter month product prodtype country region province;  
    *** la tabla nway tiene 100,000 celdas;  
  var sales/sum;  
  hierarchy year quarter month/display=yes;  
  hierarchy prodtype product/ display=yes;  
run;
```

Una solicitud realizada en el visor indica que se quiere ver el total de ventas (sales) de cada país (country). Existen 12 ciudades, así que una subtabla con 12 celdas sería la óptima, sin embargo no se encuentra. Si la subtabla óptima no se encuentra disponible, entonces se usará una tabla que contenga la variable country. En este caso, la única tabla disponible que contiene la variable country es la tabla NWAY, la cual contiene 100,000 celdas.

Los siguientes pasos se realizan al hacer la solicitud:

El modelo hace una solicitud de datos al servidor. En este caso la suma de las ventas (sales) de los diferentes países (Country).

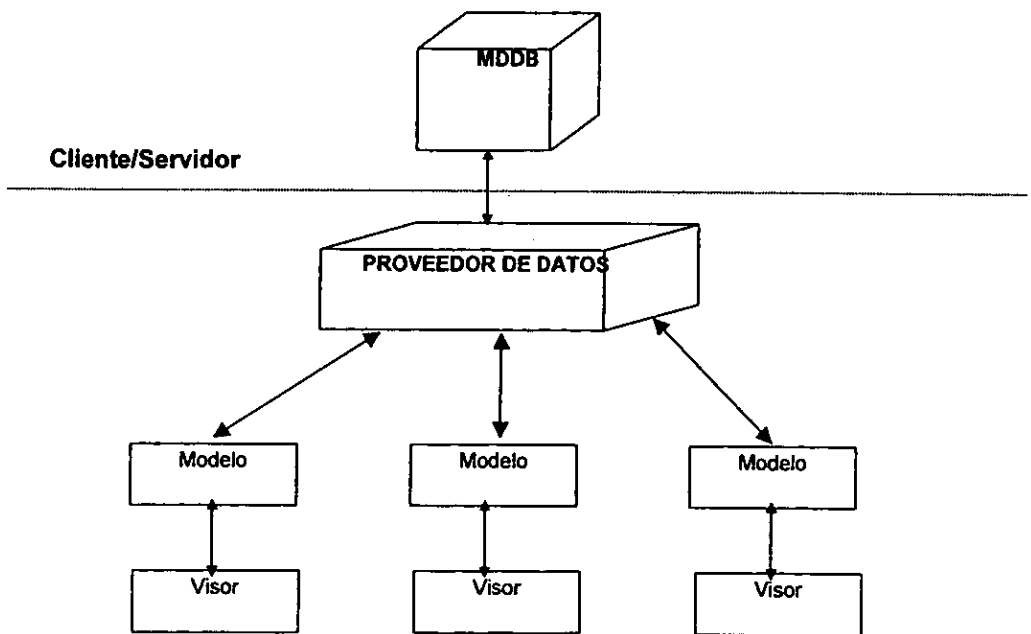
- En el servidor se selecciona del MDDB la subtabla más apropiada. En este caso la tabla NWAY.
- Toda la tabla es transferida del servidor a la memoria en disco del cliente. Mueve 100,000 celdas de información.



- La sumalización para obtener la subtabla que se requirió es realizada en el cliente (sólo si es necesario). En este caso de 100,000 celdas, se deja la sumalización dentro de las 12 celdas requeridas (el número de países).
- La información requerida es desplegada en el visor. Es decir, despliega las ventas por cada país.

Como se puede ver, esto se vuelve un problema, ya que cuando se hace una solicitud de una subtabla pequeña y no se encuentra, se tendrá que traer la tabla NWAY, y sería muy problemático si esta tabla es muy grande.

Ahora consideremos la misma situación, pero usando Holap, esto se vería como lo muestra la figura siguiente:



Estructura con HOLAP



Al igual que el caso anterior, el visor sabe lo que deberá desplegar al usuario. Lo mismo sucede con las fases que se encuentran entre el visor y la fuente de datos.

El Proveedor de datos se encuentra en el cliente entre el (los) modelo(s) y la fuente de datos. Entre otras cosas sabe:

- Recibir solicitudes del modelo
- Manejo al acceso de datos

Con el HOLAP, para el ejemplo anterior se generan los siguientes pasos:

- El modelo hace una solicitud de datos al proveedor de datos. En el caso la suma de las ventas (sales) de los diferentes países (country).
- El proveedor de datos realiza una tarea remota a la fuente de datos apropiados (sólo si es necesario, es decir si no se encuentra disponible la subtabla que cumpla con los requisitos de la solicitud). En este realiza la tarea de elegir la tabla NWAY.
- Cualquier sumariación y subconjunto de datos se realiza en el servidor. En este caso se crea un MDDDB con 12 celdas, donde los datos para crear este pequeño MDDDB se toman de la tabla NWAY elegida anteriormente.
- Sólo el resultado de la consulta es bajado a la máquina del cliente, de tal forma que ese resultado lo presente el visor.

Como se puede ver, en el HOLAP las grandes transferencia de datos no son necesarias, ya que sólo se requerirá la transferencia del resultado, puesto que todos los procesos (cálculos y generación de subtablas) son realizados en el servidor.

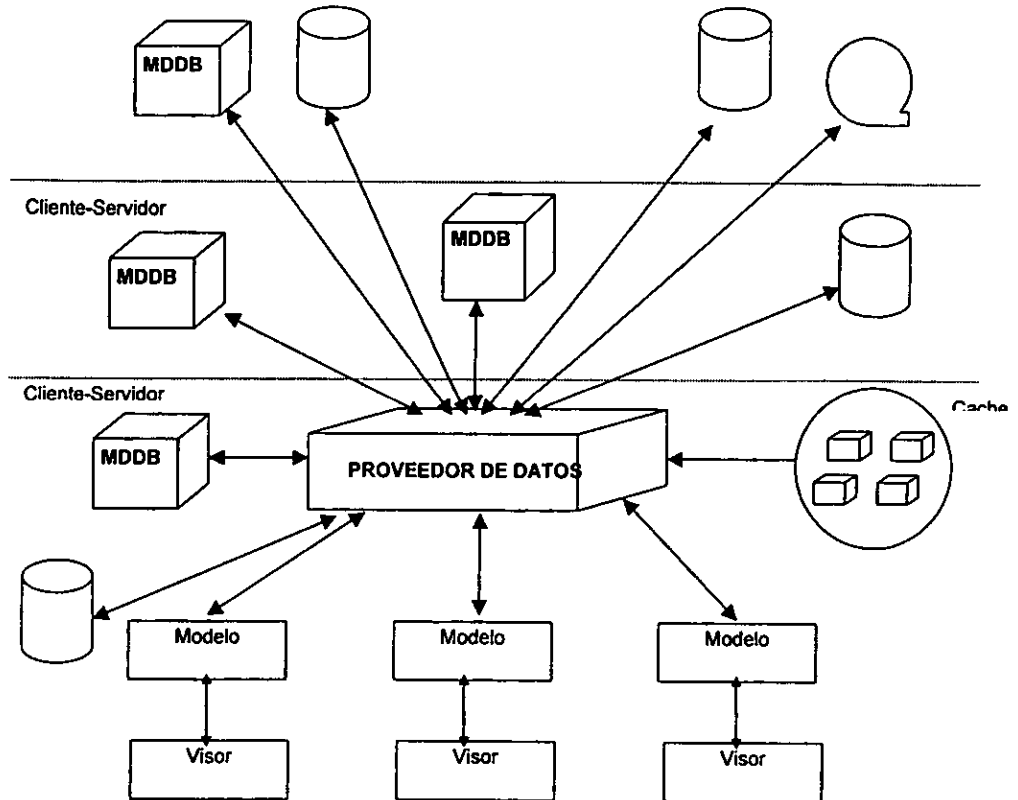
El resultado de una consulta con HOLAP siempre será un MDDDB que contendrá la información que satisfaga la consulta, es decir, un MDDDB que contendrá





únicamente las celdas que se requieran. Este MDDB es bajado al cliente para desplegar el resultado, además de mantenerse en cache, los elementos del cache están disponibles para ser usados en futuras consultas.

HOLAP también puede ser usado para acceder a un gran número de fuentes de datos diferentes, tal como lo muestra la siguiente figura:



Acceso a diferentes fuentes de datos con HOLAP



Sólo existe un proveedor de datos con un cache. El proveedor de datos puede acceder diferentes fuentes de datos que probablemente residan locales o en uno ó más servidores.

Una de las características importantes del HOLAP es que tiene la habilidad de realizar cálculos y procesos en el servidor. El proveedor de datos genera el código localmente cuando se realiza una solicitud. El código se ejecuta remoto y el resultado de la solicitud es depositado en un MDDB, el cual es bajado al cliente para desplegar el resultado.

Las situaciones por las cuales se realizan los cálculos y procesos en el server:

La fuente de datos no es un MDDB.

La fuente de datos es un MDDB en el cliente, pero las variables de clase no concuerdan exactamente.

Algunas veces el funcionamiento de generación de cálculos en el servidor del MDDB es menos eficiente que simplemente una lectura via RLS (acceso de datos), esto se debe a que cuando se genera alguna consulta se tiene la tabla requerida, sin embargo ya se generó el código, o la subtabla por la que se tomarán los datos no sea tan grande y el acceso a la red sea muy lento.



## Reflexión de la herramienta SAS

Gracias al estudio que se realizó de la herramienta SAS se pudo hacer un mejor diseño en lo que fue las bases de datos multidimensionales, ya que la primera forma en que estos se realizaron fue errónea, los cubos que se diseñaron fueron uno por mes para cada uno de los reportes que se tenían, y por las diferentes tipos de dimensiones, de tal forma que para el reporte de facturación se tenían 39. Esto por el número de meses (trece) que se tenía de información y para cada forma de navegación que se tuviera, es decir, las tres diferentes dimensiones que se dieron (por estado, por agrupación y por cadena).

Lo mismo pasaba para el consolidado anual, salvo que para este sólo eran tres cubos, pues en este se tenían todos los meses. Para este consolidado no se llegó a nivel de comercio, ya que el tamaño que podía tomar este superaba los 2 Gbytes, para lo cual el Sistema Operativo (solaris) no acepta almacenamiento mayor a 2 GB.

Para análisis libre sólo se requirió uno por mes, puesto que en este el usuario podría crear su propia consulta (inicialmente no se llegó a nivel de comercio).

La única diferencia que existía entre estos era la variable de análisis, donde esta no afecta mucho el crecimiento del cubo. Fue entonces que se decidió crear un solo cubo por mes para todos los tipos de reporte, en este se incluirían todas las variables de análisis que se requirieran.

También se implementó la funcionalidad de HOLAP, puesto que la aplicación al ser instalada a los usuarios (bancos) se mostraba con un rendimiento muy pobre, esto se debió a que las tablas que se requerían en los reportes al consultar un cubo tenían que pasar por la red, así que si se tenía mucho tráfico de red y además un canal de comunicación muy estrecho, implicaba que el tiempo de



respuesta llegara a ser hasta de 30 minutos, lo cual no era operable. Con la técnica HOLAP, esto se redujo a un tiempo de respuesta de hasta 15 segundos, se debe aclarar que este tiempo de respuesta es muy razonable por la cantidad de información que se maneja y la forma en que se presenta.

Además, al incluir la técnica HOLAP, en el reporte de análisis libre se pudo llegar a un nivel de detalle de comercios y los cubos que inicialmente llegaron a ser 39 se redujeron a 13, es decir, uno por mes, ya que se dejaba que algunas sumalizaciones se realizaran en el servidor tal y como lo maneja la técnica HOLAP, de esta forma no se tendrían que realizar demasiadas subtablas para un cubo.



# Capítulo VII.

## Diseño del sistema

**Todo arte es sólo la imitación de la naturaleza.**

**Séneca**

**La imaginación abandonada de la razón crea monstruosidades. Unida con la razón, la imaginación da lugar a grandes maravillas y a verdadero arte.**

**Goya**



El módulo adquirente fue diseñado con la finalidad de proporcionar información a los <bancos del consorcio> generando así reportes de sus facturaciones, transacciones, comercios activos presentes y la rentabilidad que tienen.

Éste módulo se encuentra en el Sistema de Información estratégica de la empresa. Para la generación de estos reportes, el sistema se encuentra desarrollado en SAS y pretende agilizar la generación de información. El cliente final podrá realizar análisis no definido con los datos que se le presentan, el cliente se conectará directamente a la base de datos desde su PC.

Se presentará el análisis estructurado del sistema, el cual contiene la lista de eventos que actúan en él, procesos que en éste se realizan para responder a dichos eventos, tablas de información utilizadas en el módulo, estructura de tablas finales e intermedias que se utilizan.

El propósito del sistema de información estratégico Módulo Adquirente es emitir reportes y pronósticos de las facturaciones, transacciones, los comercios que se encuentran activos en el mes presentado y la rentabilidad de los bancos para con los comercios, esto con el fin de facilitar la administración de información, basándose en análisis estadístico para una ágil toma de decisiones.

También incluye generar una base de datos reducida para contener solamente la información estratégica, mantener el repositorio de datos actualizado, emitir reportes por banco adquirente y presentación de gráficas de los datos.



## Lista de eventos

1. El departamento de Proyectos de la empresa requiere consultar información mensual.
2. El departamento de Proyectos solicita reporte mensual de facturación, transacciones, comercios activos, rentabilidad, ingresos y ticket promedio mostrado por Banco - población o cadena - Agrupación y giro.
3. El departamento de Proyectos solicita reporte de pronósticos de las facturaciones y transacciones en 13 meses.
4. Banco del consorcio solicita reporte mensual de facturación, transacciones, comercios activos, rentabilidad y ticket promedio mostrado por población o cadena - Agrupación, giro y razón social del comercio.
5. Banco del consorcio solicita reporte de pronósticos de las facturaciones y transacciones en 13 meses.

## Diagrama de Contexto.

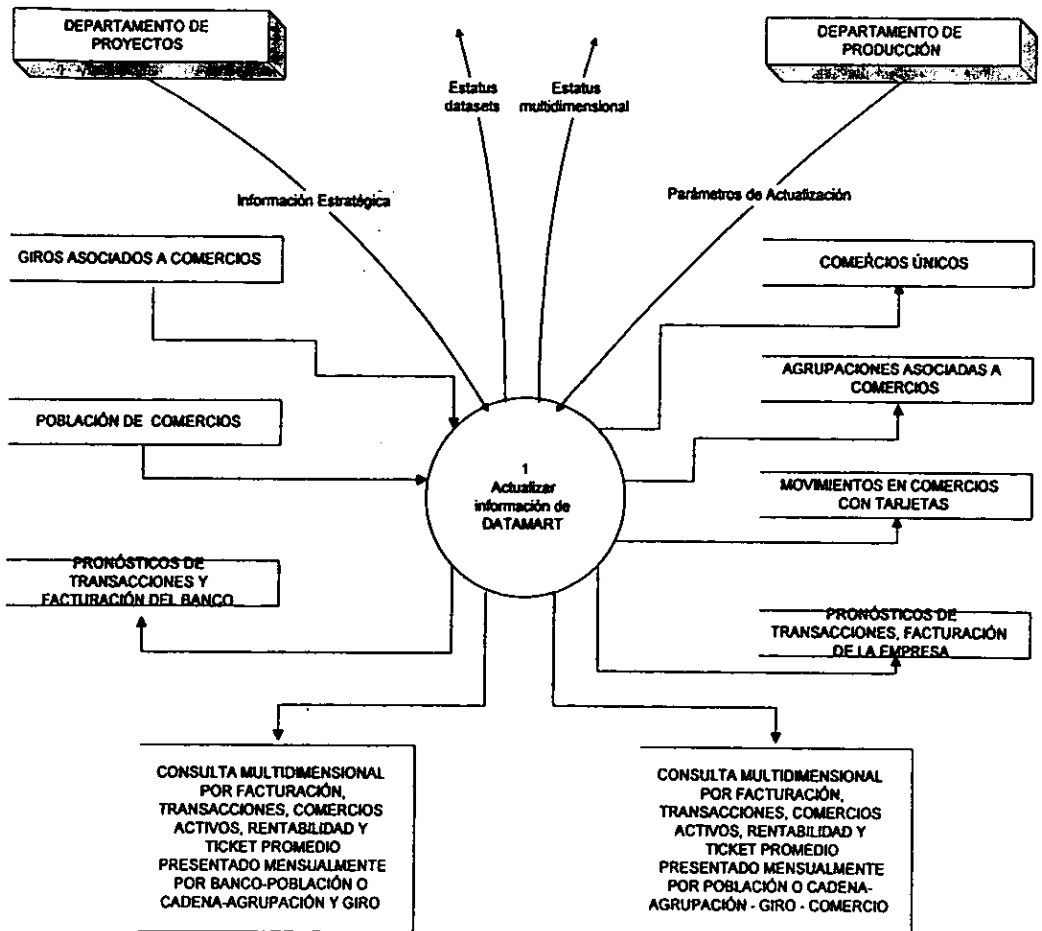




## Diagrama de Flujo de Datos (DFD).

### Descripción de eventos

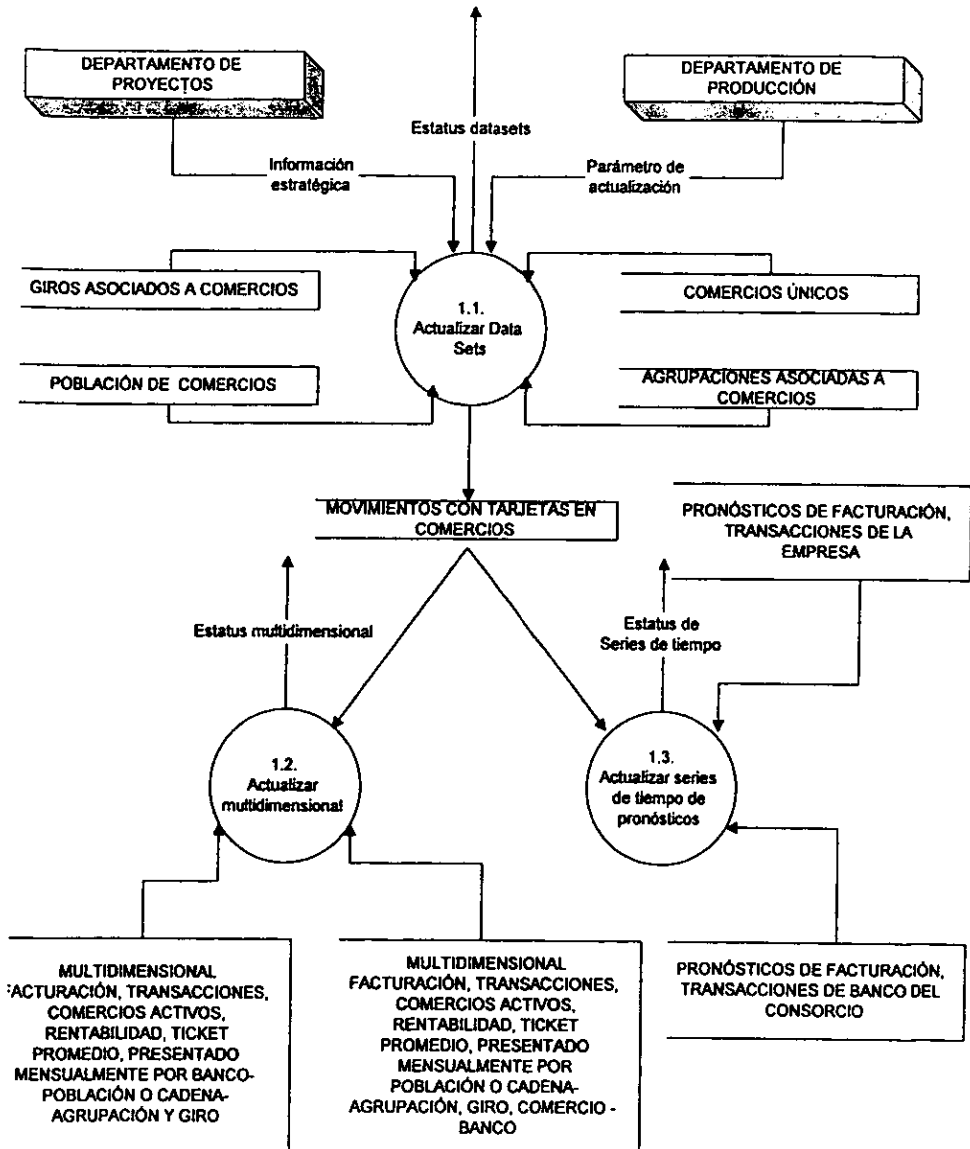
Evento 1: El departamento de información estratégica requiere consultar información mensual.





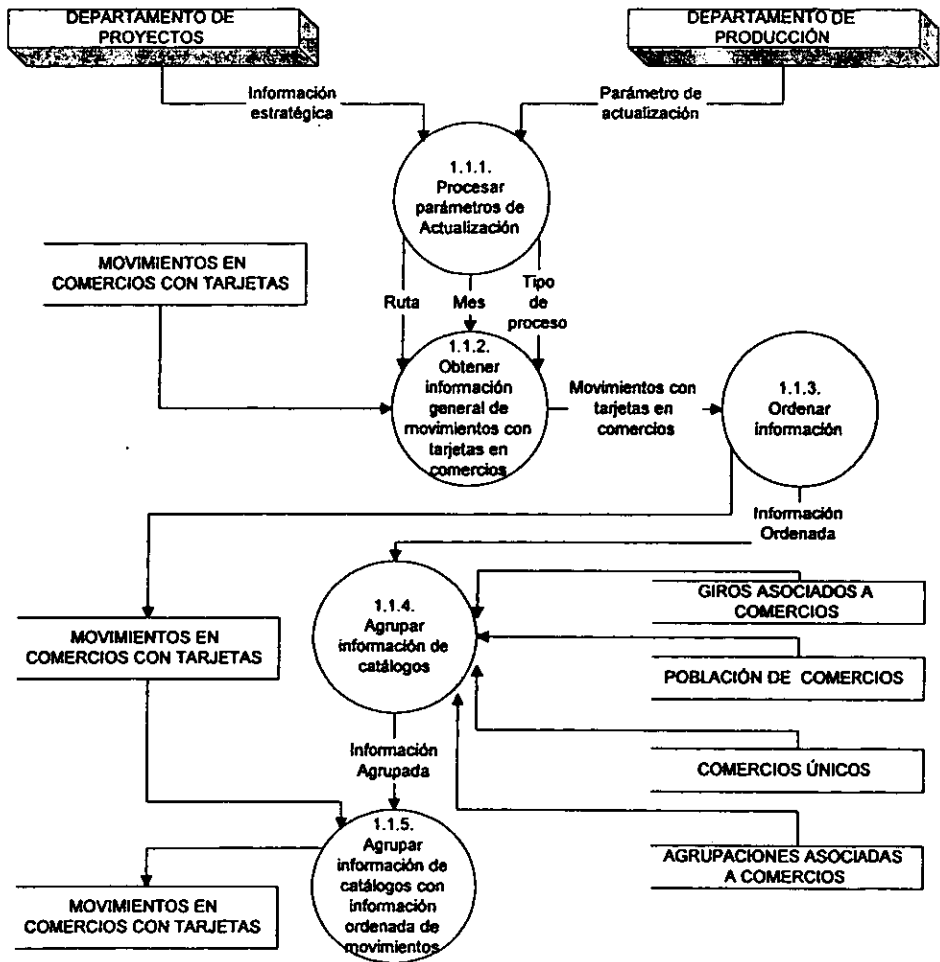


### Nivel 1



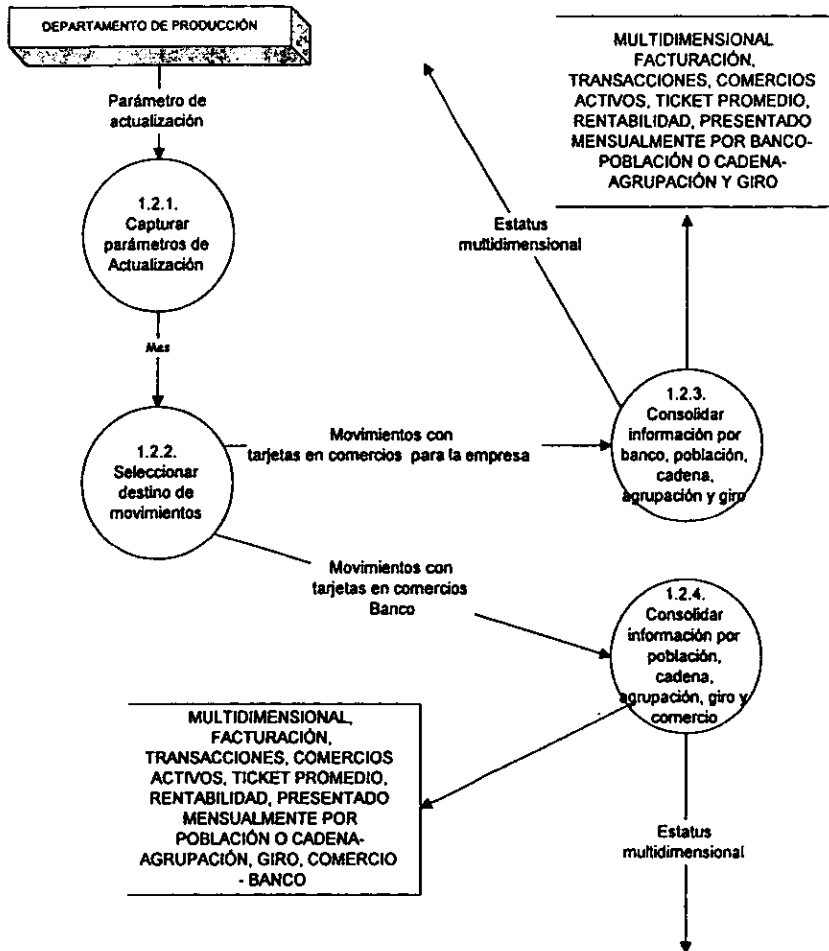


## Nivel 2



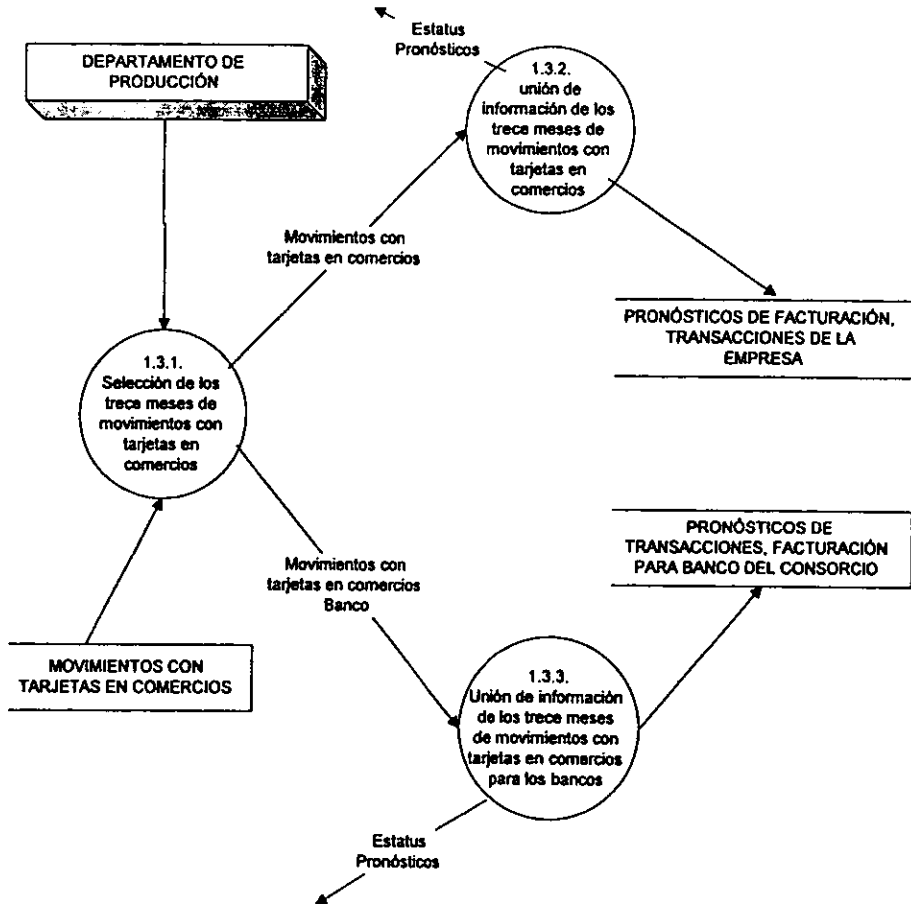


## Nivel 2



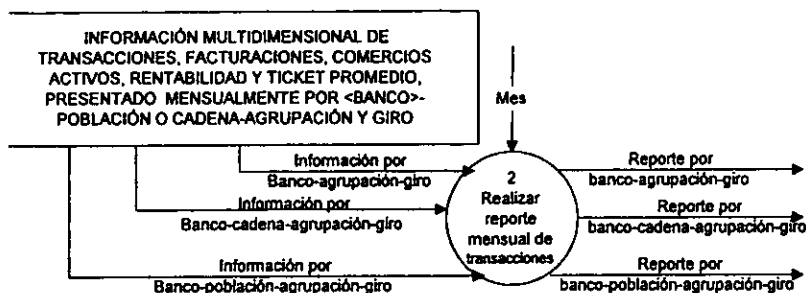


## Nivel 2

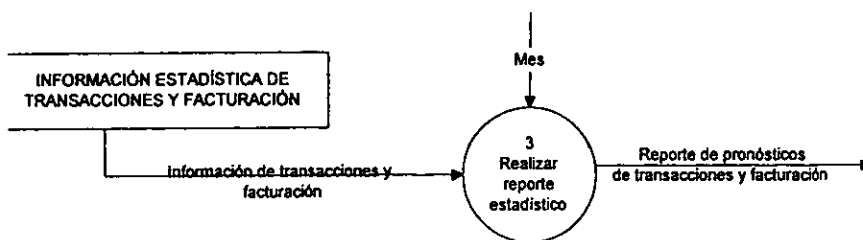




**Evento 2:** Departamento de proyectos estratégicos solicita reporte mensual de facturación, transacciones, comercios activos, rentabilidad, ingresos, ticket promedio, lo cual se pide sea mostrado por banco - población o cadena - Agrupación y giro.

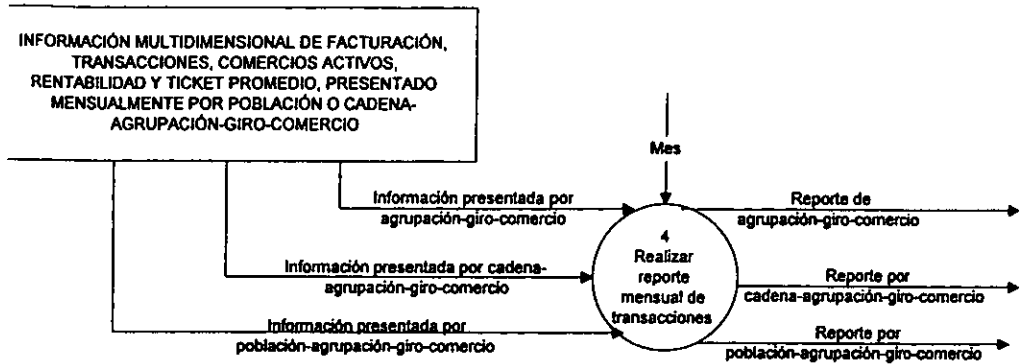


**Evento 3:** El departamento de proyectos estratégicos solicita reporte de pronósticos de las facturaciones y transacciones en 13 meses.

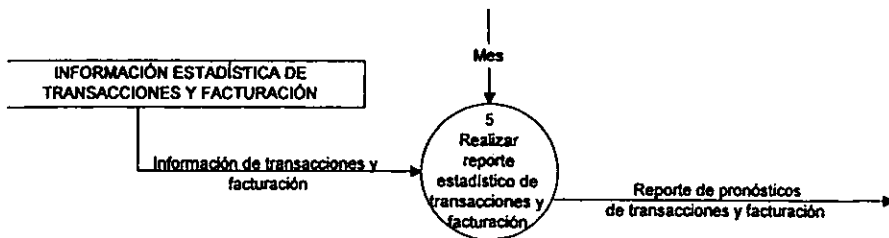




**Evento 4:** Banco del consorcio solicita reporte mensual de facturación, transacciones, comercios activos, rentabilidad, ingresos, ticket promedio, lo cual se pide sea mostrado por: Población o cadena - Agrupación, giro y razón social del comercio.



**Evento 5:** Banco del consorcio solicita reporte de pronósticos de las facturaciones y transacciones en 13 meses.





## ANÁLISIS DE TABLAS

Giro.

Catálogo de giros, contiene también agrupación para comercios, ésta tabla se deposita en el servidor y se podrá modificar por medio de pantallas, contiene las agrupaciones, se pone formato a la agrupación, ya que su descripción es muy grande. Este archivo se encuentra en formato SAS.

Campo	Longitud SAS	Formato	Descripción
Gir_agru	\$1	<Agrupa>*	Agrupación del giro
Gir_cve	4	Integer 9.	Clave del giro
Descrip	\$35	Char 35.	Descripción del giro
Giro_sic	8	Integer 10.	Clave internacional

\*<Agrupa>:

'H' = 'Hoteles'

'R' = 'Restaurantes'

'L' = 'Línea aéreas'

'T' = 'Tiendas'

'O' = 'Otros'

**other**= 'Desconocido';



### Cadenas.

Contiene Información de las cadenas que existen, se deposita como archivo ascii en el servidor. Y se convierte a formato SAS. Este archivo se genera cada semana.

Campo	Longitud SAS	Formato	Descripción
Cad_gpo	4	Integer 4	Número del Grupo de cadena
Cad_cve	4	Integer 4.	Clave de la cadena
Nom_cad	\$36	char 4	Nombre de la cadena
Nkey	4	Integer 4.	Llave para la cadena

### Poblaci1.

Contiene información de las Poblaciones y su clave, éste archivo ascii se deposita en el servidor y se convierte a formato SAS, se pretende poner formato al estado. Este archivo se genera cada semana.

Campo	Longitud SAS	Formato	Descripción
Cve_pobl	4	Integer 4.	Clave de la población
Cve_edo	4	<Estado>	Clave del estado
Nom_pobl	\$36	Integer 4.	Población

### Bducomer.

Contiene información de los Comercios que se encuentran dados de alta en el archivo único de comercios, este archivo se deposita en el servidor en formato ascii, y se convierte a formato SAS se le pone formato al comercio, ya que su descripción es muy grande. Este archivo se genera cada semana.





Campo	Longitud SAS	Formato	Descripción
Comfilia	6	Integer 6	Número de afiliación del comercio
Gir_cve	4	Integer 4	Giro al que pertenecen
Com_Nomb	\$35	Char 35	Nombre del comercio
Cad_Cve	4	Integer 4	Clave de la cadena
Cad_Gpo	4	Integer 4	Grupo de la cadena
Bdu_Intr	3		Categoría del comercio (no se ocupa)
Com_Cate	8		Categoría de débito
Cve_pob	8		Clave de la Población del comercio

**Intred.**

Contiene información de las Interredes, este archivo será modificado cada vez que se de una intered de alta, se podrá modificar a través de pantallas.

Campo	Longitud SAS	Formato	Descripción
Intred_g	8	Integer 8	Clave de la intered
D_intred	\$20	Chart 20.	Cadena

**Tabla1.**

Resultado de la extracción del Archivo ascii 136, este archivo es el que se genera en el momento en que se realizan movimientos en las terminales POS. Contiene la información de los movimientos con tarjetas en comercios, lleva ciertas restricciones (ver miniespecificaciones). La extracción de este archivo se deja como temporal, una vez generada la tabla final se elimina.



Campo	Longitud SAS	Formato	Descripción
Transacc	3	Transac.	Tipo de transacción
Emisor	3	Bancos.	Banco Emisor
Receptor	3	Bancos.	Banco Receptor
Cuenta	8	19.	Número de Cuenta
Importe	8	13.2	Importe de la transacción
Nautori	\$6	\$6.	Número de autorización
Fproceso	8	DDmmyy8.	Fecha de proceso
Fconsumo	8	DDmmyy8.	Fecha de consumo
Comfilia	8	Comercio.	Clave del comercio
Stransac	\$1	\$Stran.	Situación de la transacción
Icredeb	\$1	\$Credebi.	Identificación de Crédito Débito
Ibinemit	\$1	\$Bin.	Identificación del Bin emitido
Numfuen	3		Número de fuente
Procesad	\$1	\$Procesa	Procesador Administrador
Importec	8	10.2	Importe de la comisión
Ivacomi	8	7.2	Importe del iva de la comisión
Itarcomi	\$1	\$Tarifco.	Indicador de Tarifa o comisión
Inointer	\$1	\$Intered	Indicador de terifa o comisión
Tdbanemi	8	10.2	Costo de Intercambio
Mentrada	\$2	\$Entrada.	Modo de Entrada
Intred_g	3	Intred.	Intered
Propio	3	propio	Procesado por
Factura	8	13.2	Facturación (Importe-devoluciones)
Semana	3		Semana de consumo
Date	\$5	Date	Año/mes



TablaF.

Resultado de la relación de las tablas maestro con la tabla1, contiene todos los movimientos de las tarjetas, el comercio al que pertenecen, el banco adquirente que corresponden, etc.

Campo	Longitud SAS	Formato	Descripción
Transacc	3	Transac.	Tipo de transacción
Emisor	3	Bancos.	Banco Emisor
Receptor	3	Bancos.	Banco Receptor
Cuenta	8	Integer 19.	Número de Cuenta
Importe	8	Flotante 13.2	Importe de la transacción
Nautori	\$6	Char 6.	Número de autorización
Fproceso	8	DDmmyy8.	Fecha de proceso
Fconsumo	8	DDmmyy8.	Fecha de consumo
Comfilia	8	Comercio.	Clave del comercio
Stransac	\$1	\$Stran.	Situación de la transacción
Icredeb	\$1	\$Credebi.	Identificación de Crédito Débito
Ibinemit	\$1	\$Bin.	Identificación del Bin emitido
Numfuen	3	Integer 3	Número de fuente
Procesad	\$1	\$Procesa	Procesador Administrador
Importec	8	Flotante 10.2	Importe de la comisión
Ivacomi	8	Flotante 7.2	Importe del iva de la comisión
Itarcomi	\$1	\$Tarifco.	Indicador de Tarifa o comisión
Inointer	\$1	\$Intered	Indicador de terifa o comisión
Tdbanemi	8	Flotante 10.2	Costo de Intercambio
Mentrada	\$2	\$Entrada.	Modo de Entrada
Fech_cga	3	Date7.	Fecha de Carga
Propio	3	Propio.	Tipo de proceso (on us, us off)
Factura	8	Flotante13.2	Facturación
Semana	3	Integer 3	Semana de consumo
Date	\$5	Chart 5	Año/mes



Cve_acad	4	Integer 4	Agrupación de cadena comercial
Cve_cade	4	Cadena.	Nombre de cadena Comercial
Intred_g	3	Intred.	Grupo de Intered
Gir_cve	4	Giros.	Giro comercial
Com_cate	3	Integer 3	Comisión de intercambio
Cve_Pobl	4	Poblaci.	Poblado
Cve_edo	4	Estado.	Estado
Gir_agru	\$1	\$Agrupa.	Agrupación de giro internacional
Num_come	3	Integer 3	Número de comercios Activos
Numt	6	Integer 6	Número de transacciones



# Capítulo VIII.

## Conclusiones

**Todo el bien permanece con aquel que  
esperó sabiamente.**

**Thoreau**



Hoy en día, existe una gran competencia entre las empresas, ahora todo mundo tiene acceso a los datos, pero lo más importante es que tan rápido y eficiente sean procesados para ser convertidos en información para la toma de decisiones.

La gran mayoría de las empresas se encuentran con sus sistemas operacionales funcionando al 100% y es ahora que empiezan a darse cuenta de la necesidad de tener sistemas de información que no sólo contengan los históricos de la operación de la empresa, sino también que sean una herramienta para la toma de decisiones.

Es por eso que las empresas se preocupan por diseñar y mejorar sus sistemas de información, teniendo sistemas de datos centralizados como serían los modelos de datos de Data Warehouse evitando incongruencias y duplicidad de los datos que serán incorporados a los sistemas de toma de decisión.

Creando un repositorio de datos (DataWarehouse) se tiene una mejor visión del negocio accediendo eficientemente cualquier tema de interés, de esta manera los directivos tienen en sus manos cualquier información relacionada a sus empresa, actualizada y fácil de consultar.

Las empresas dedicadas al ámbito financiero, como la que se describe en este trabajo requieren de sistemas de información que les proporcionen una mejor visión en el negocio, teniendo rápido acceso a la información, además de poder manipularla como mejor les convenga.

Las aplicaciones de procesos en línea (OLAP) están proveiendo la capacidad de analizar información visualizando las relaciones existentes entre una y otra variable.



Las empresas dedicadas a generar productos que proporcionen solución a los problemas de información se están preocupando más por realizar herramientas que proporcionen un rápido acceso a los datos y manejo de los mismos, así como el rápido almacenamiento y una capacidad de espacio cada vez mayor.

Los beneficios que se obtienen al tener un sistema de información estratégico son muchos, como muestra del generado para la empresa mencionada en este trabajo se tienen:

- No se requiere de mucho personal para la generación de reportes.
- Los reportes pueden generarse en un tiempo mínimo, ya no es necesario que el usuario (banco del del consorcio) tenga que esperar días para una respuesta a su petición, ahora requiere de unos cuantos segundos para poder visualizar su información y con ello detectar anomalías por parte de las tiendas asociadas a el.
- El usuario puede dar el formato que desee a su reporte.
- Se puede tener un historico de información a 13 meses hacia atrás a partir de la fecha de consulta.
- Se puede realizar un reporte generando diferentes combinaciones, es decir se pueden encontrar relación con diferentes variables, sin necesidad de que se tenga que sacrificar tiempo.
- Se tiene mejor visualización del negocio.
- No existen descontentos por parte del usuario, ya que la información que consulta es clara y confiable.
- El usuario puede visualizar mejor su negocio, es decir visualizar los comercios que le están generando pérdidas, así como aquellos que están generando las mayores ganancias, y con ello implementar un método para mantener un mejor control.



# BIBLIOGRAFIA





- Publication Catalog  
SAS Institute Inc.
- SAS/Warehouse administrator. Metadata API Reference  
SAS Institute Inc.
- SAS/Warehouse administrator. User's Guide, Release 1.1  
SAS Institute Inc.  
First Edition
- El libro blanco de SAS  
SAS Institute Inc.
- SAS/EIS  
First Edition
- SAS MDDDB server  
First Edition

#### Internet

- [www.sas.com](http://www.sas.com)



# ANEXO



## Interfaz del Sistema de Información Estratégica

La interfaz del Sistema es la siguiente:

La forma como se ve la información es por facturación, transacciones, comercios activos, rentabilidad (importe de la comisión y costo de intercambio) y una parte de análisis libre, en la cual se puede manipular la información de la forma que se requiera.

Toda la información se puede visualizar de tres formas:

**Agrupación:** La información se consulta por

Agrupación

Giro

Comercio

**Estado:** La información se consulta por

Estado

Agrupación

Giro

Comercio

**Cadena:** La información se consulta por

Cadena

Agrupación

Giro

Comercio

Inicialmente, se tiene un folder que presenta 4 pestañas, éstas muestran las

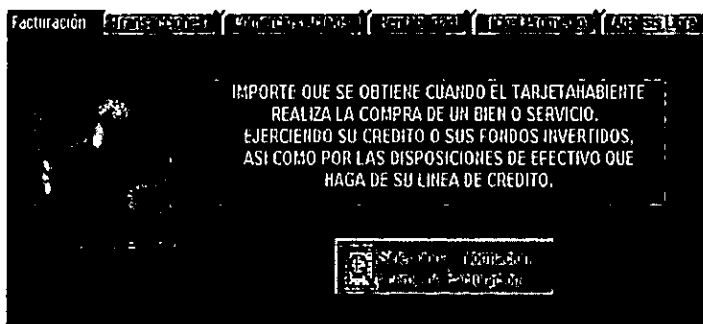


diferentes opciones de consulta que tiene el sistema, sólo se describirá la parte de Adquirente.



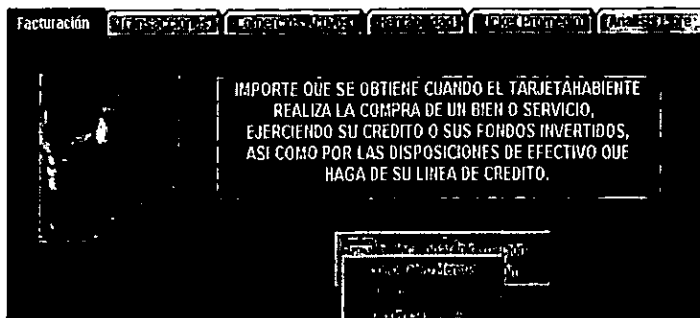
Opciones del Sistema

Dentro de Adquirente, indican seis pestañas, a través de las cuales se puede acceder a las diferentes opciones de consulta:



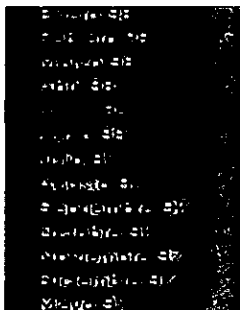
Opciones de consulta del módulo adquirente

El sistema tiene la opción de visualizar un comparativo mensual (ver 13 meses), información por mes y una tendencia de la información que se nos está presentando.



Diferentes dimensiones para visualizar la información

Si se desea ver la información mensual, se debe elegir el mes a analizar, por ejemplo mayo 98. Esta lista de meses dependerá de la fecha en que se este accediendo al sistema, ya que este presentará información de 13 mese hacia atrás a partir de la fecha de consulta.



Lista de meses a consultar, depende de la fecha de consulta

Una vez seleccionado el mes a consultar se presenta una pantalla que contiene la información que se está solicitando, esta es la siguiente:



Forma en la que se está  
 presentando la  
 información

Envía Información  
 a Excel

Imprime reporte

Presenta  
 Información  
 en forma  
 gráfica

The screenshot shows a window titled 'Facturación' with a menu bar containing 'Gráficas', 'Reporte', and 'AGREGACIÓN'. Below the menu is a 'Mostrar' dropdown set to 'ALL'. The main area contains a table with columns for 'Tipo de transacción', 'Monto', and 'ACTUAL'. The table lists various transactions with their respective amounts. Below the table, it says 'Información solicitada'. To the right of the table is a control panel with several dropdown menus: 'Monto de consulta', 'Mantención de la transacción', 'Tipo de operación', 'Fecha', 'Tipo de periodo', and 'Intervalo', all set to 'ALL'. At the bottom, a message states: 'Esta información no contiene el monto: Disposiciones en Ventanilla.' There are also buttons for 'Excel' and 'Imprimir'.

Tipo de transacción	Monto	ACTUAL
Apertura de Cuenta	37,000,000.00	37,000,000.00
Depositos	113,041,054.00	9,853,367.72
Retiros	91,30,000,000.00	5,114,000,000.00
Otros	21,70,000,000.00	5,070,000,000.00
Transferencias	117,537,100.00	6,000,000.00
Total	90,000,000.00	6,300,000.00

Diferentes  
 formas de  
 consulta

Opciones  
 para  
 condicionar la  
 información

Información solicitada

El formato contenido para las consultas de facturación, transacciones, comercios activos y rentabilidad es similar al presentado en la pantalla anterior.

Para la consulta de análisis libre, el usuario tiene la opción de consultar su información como él considere conveniente de acuerdo a sus necesidades.

En análisis libre se cuenta con un botón llamado VISTA, el cual permite seleccionar la información que se requiera.

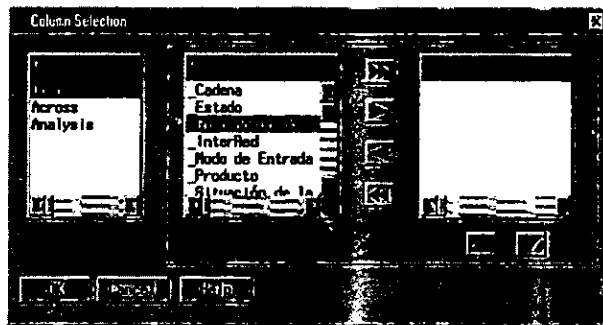


Botón de vista

Cadena	Estado	Interfied	Modo de Entrada	Producto	Producción de la
506	\$208,777.15	\$14,270.95	\$15,226.30		
512	\$247,223.57	\$4,429.90	\$8,272.51		
111	\$0,525.86	\$0.00	\$0.00		
105,248	\$153,187,762.67	\$3,848,187.80	\$3,281,787.87		
60,747	\$75,843,743.08	\$1,840,754.04	\$1,554,501.87		
09	\$217,891.36	\$0.00	\$0.00		
3	\$1,788.89	\$0.00	\$0.00		
178	\$127,541.28	\$3,837.60	\$2,887.81		
93	\$80,263.34	\$2,307.81	\$1,835.38		
487	\$518,289.70	\$19,224.78	\$15,968.57		
278	\$219,341.85	\$3,073.35	\$6,780.59		
1,732	\$1,287,443.75	\$38,638.80	\$28,333.40		
1,326	\$864,874.82	\$27,828.14	\$20,384.21		
1,783	\$573,884.44	\$24,869.27	\$18,586.84		
1,282	\$345,263.44	\$13,517.37	\$10,017.71		
14,487	\$13,885,324.48	\$432,246.78	\$332,253.43		
10,487	\$7,877,557.09	\$224,285.88	\$174,671.88		

Primer información presentada en análisis libre

Al seleccionar la opción vista, se presenta una pantalla que cuenta con las opciones de: Across, Down y Analysis.



Pantalla generada al entrar en la opción vista



La opción Down, permite presentar la información que se desea ver por renglones (se debe seleccionar por lo menos una variable).

Esta información puede ser: **\_Modo de entrada, \_Producto, \_Situación de la transacción, \_Tipo de operación, \_Tipo de producto, \_Intered, \_Giro Agrupado, \_Transacción (Ventas y devoluciones).**

La opción de Across, indicará la información que se desea ver por columnas (se puede o no seleccionar la variable).

Esta información puede ser: **\_Modo de entrada, \_Producto, \_Situación de la transacción, \_Tipo de operación, \_Tipo de producto, \_Intered, \_Giro Agrupado, \_Transacción (Ventas y devoluciones).**

La opción de Analsys indica la forma en la que será analizada la información, por ejemplo, si se desea analizar la información de los productos (se refiere a la tarjeta, esta puede ser Visa o Mastercard), esta se puede ver por facturación, transacciones, número de comercios, costo de intercambio, tasa de descuento.

Una vez seleccionada la información que se desea analizar, al presionar el botón de OK, el resultado se presenta en un tiempo aproximado de 10 segundos.

El sistema tiene diferentes opciones para generar un reporte de acuerdo a las necesidades del usuario. Dichas opciones son presentadas al presionar botón derecho del mouse, estas se presentan en forma de lista, tal como se muestra:





Up	Sube un nivel
Down	Baja un nivel en general
Filter Down	Baja un nivel filtrando por valor seleccionado
Expand	Muestra el siguiente nivel expandido a la derecha
Contract	Si se expandió un nivel realiza lo contrario
Left	Muestra información de un campo arriba del seccionado
Right	Muestra información de un campo abajo del seleccionado
Rotate	Cambia rangones por columnas y viceversa
Report	Permite seleccionar las variables que se requieran analizar
Chart	Permite cambiar el tipo de estadística
Format	Permite maquillar el reporte (poner formatos, colores, etc.)
Computed	Permite crear una variable o campo calculado
Sub-report	Permite generar un reporte condicionado
Show Detail	Muestra el detalle del nivel en el que se encuentra el reporte
Show Summary	Muestra el nivel en que me encuentro