

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES
DE MONTERREY**



TESIS

**NIVEL DE IMPLEMENTACIÓN DE LA TECNOLOGÍA DATA WAREHOUSE EN
GRANDES EMPRESAS.**

**MAESTRÍA EN ADMINISTRACIÓN DE
TECNOLOGÍAS DE INFORMACIÓN**

POR:

ERNESTO TREVIÑO GARZA

Abril del 2002

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

**DIVISIÓN DE ELECTRÓNICA, COMPUTACIÓN, INFORMACIÓN Y
COMUNICACIONES**

**PROGRAMA DE GRADUADOS EN ELECTRÓNICA, COMPUTACIÓN,
INFORMACIÓN Y COMUNICACIONES**

Los miembros del comité de tesis, recomendamos que esta tesis presentada por Ernesto Treviño Garza sea aceptada como requisito parcial para obtener el grado académico de Maestro en Administración de Tecnologías de Información.

Comité de tesis:

**Dr. David Ángel Alanís Dávila
ASESOR PRINCIPAL**

**Lic. Alicia Astrid Romero García
SINODAL**

**Lic. Carlos Rodríguez Maillard
SINODAL**

**David A. Garza Salazar, PhD.
Director de los Programas de Postgrado en Electrónica,
Computación, Información y Comunicaciones.**

**NIVEL DE IMPLEMENTACIÓN DE LA TECNOLOGÍA DATA
WAREHOUSE EN GRANDES EMPRESAS.**



TESIS

POR:

ERNESTO TREVIÑO GARZA

**Presentada a la División de Electrónica, Computación,
Información y Comunicaciones**

**Este trabajo es requisito parcial para obtener el grado de
Maestro en Administración de Tecnologías de
Información**

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES
DE MONTERREY**

Abril del 2002

DEDICATORIA

A mi madre...

Esta tesis y en general este grado, es para ti mamá, por todo el amor y comprensión que siempre me has dado, por tus inagotables palabras cariñosas y de ánimo. Eres mucho más que la persona que me trajo al mundo, eres mi mejor amiga, mi soporte en momentos difíciles y una de las dos personas que más amo.

A mi padre...

Pelao, por ser más que una guía en mi camino, mi máximo apoyo en cada proyecto que inicio, por enseñarme el valor de la familia. Por tu amor y comprensión en todo momento, sin importar el camino que tome. Por todos tus "Estoy muy orgulloso de ti" que han sido mi mayor alimento en la lucha por mis metas. Gracias a ti soy el hombre que hoy conoces. Por ser la otra persona que más amo.

A mi hermano Alberto...

Por la alegría que les diste a mis papás con su primer nieto. Por tu manera de hacerle frente a la vida sin importar lo difícil que se puedan poner las cosas. Te quiero mucho.

A mi hermano Alejandro...

Por esas charlas nocturnas por el msn que hacen que la distancia familiar no se sienta tanto, por ser el que más me demuestra su gozo con mis triunfos y me exige cada vez más. Por confiar tanto en mí. Por esas lágrimas fuera de casa de mi abue cuando me dejaron por primera vez en Mty, por eso y mucho más... Te quiero mucho.
(Saca las gírlas!!)

A mi hermano Arturo...

Por ser mi pareja favorita para el dominó, por todos los momentos de apoyo para seguirle dando a la tesis, por estar aquí y compartir conmigo tus logros, por prestarme tu lap para terminar mi tesis, por prepararme esas pechugas con spaghetti ... Te quiero mucho.

A mis hermanos escogidos...

Stevin, Luis Enrique, Kim, Giancarlo, Roman, Alejandra. Por tantos momentos inolvidables compartidos, por lo mucho que me han enseñado en todos aspectos, por su cariño y amistad sincera y por lo que nos falta recorrer...

AGRADECIMIENTOS

A mi asesor...

Dr. David A. Alanís, por dedicarme su tiempo y compartir conmigo su experiencia y conocimientos. Por todas las palabras de ánimo durante esta travesía. Gracias.

A mis sinodales...

Astrid y Carlos, por su apoyo, dedicación y tiempo. Muchas Gracias.

Al CGM...

David, Daya, Hugo, Josué, Julio, León, Lepe, Maryna. Porque sin su ayuda esta maestría hubiera sido muuuuy diferente. Muchas gracias por su amistad, por todos los momentos y recuerdos que harán de esta etapa de mi vida un capítulo inolvidable, gracias por las noches de dominó, por las presidentes, por los pasos de banda, por la diversidad cultural, por los jugos jumex, en fin, gracias simplemente por estar allí.

Al ing. José Ángel Núñez...

Por brindar el apoyo técnico para la realización de mi maestría, por su amistad con mi padre, lo cual me permite asegurar que si existen amigos para siempre. Muchas Gracias.

A mi tío Meme y mi primo Oscar...

Muchas gracias por sus comentarios de psicología inversa que ayudaron a hacer mucho más agradable mi duelo con la tesis, por hacerme sentir que puedo contar con ustedes en cualquier momento. Por su cariño, Muchas Gracias.

A mis abuelas...

Por sus palabras de aliento, por su cariño, por las comidas de los lunes, por su manera tan jovial de ver el mundo, porque ambas son la prueba inequívoca de que un Dios existe. Muchas Gracias. Las quiero mucho.

A Nato y Fer...

Por aguantarme y hacer del depto 7 un lugar agradable para vivir, por los pasos de salsa, los días de gym, de fron, los desayunos domingueros "en familia", por su apoyo moral y su gran amistad.... Muchas Gracias

Y a todos lo demás que no están aquí por razones de espacio, pero que de una u otra manera han contribuido a que vea el mundo de esta manera... Muchas Gracias.

Resumen:

Tomar mejores decisiones de negocios rápidamente es la llave para el éxito en el mercado competitivo que se vive actualmente. Comprendiendo esto, las organizaciones están buscando mejorar sus sistemas de toma de decisiones ya que pueden ser rebasados por el volumen y la complejidad de los datos disponibles provenientes de sus sistemas de producción y transaccionales. El permitir que todos estos datos estén disponibles para toda la audiencia a lo largo de la empresa es hoy por hoy uno de los retos más significativos para los profesionales involucrados en las tecnologías de información. (Microsoft 1998)

Buscando la administración de ese conocimiento tan importante en nuestros días, surge una de las opciones que se han implementado para cubrir esta necesidad de las empresas, que es lo que se conoce desde hace tiempo como *Data Warehouse* (DWH), el cual, es el proceso de extraer datos de diferentes sistemas, funciones, y lugares de la empresa, agregando, filtrando, resumiendo, estandarizando, transformando y clarificando los datos extraídos e integrarlos después en una base consolidada (Anthem Inc, 2001).

El objetivo de esta investigación es el de brindar información que permita al lector tener una idea del grado de avance que tienen las empresas grandes en la ciudad de Monterrey, con respecto a la implementación de un sistema de *data warehouse*, así como los factores críticos de éxito que deben considerarse, tomando en cuenta los motivos y problemas que enfrentaron estas empresas.

Las herramientas de *data warehouse* son un área de oportunidad para muchas de las empresas grandes de la ciudad de Monterrey, ya que solo un poco más de la mitad de las empresas cuentan con esta herramienta, a pesar de estar completamente demostrada su utilidad y capacidad.

La aplicación efectiva y eficiente de una herramienta de esta naturaleza puede marcar la diferencia entre una empresa exitosa y una del montón, ya que el contar con la información correcta, completa y oportuna es una ventaja competitiva que cada vez toma más importancia en un mundo globalizado donde la información se vuelve cada vez más un factor de diferenciación.

Data warehouse parece una buena opción para aquellas empresas con un gran volumen de datos que necesitan contar con administración y acceso rápido y digerido de la información que sus sistemas transaccionales generan.

Tabla de contenidos

Dedicatorias

Agradecimientos

Resumen

Tabla de contenidos

Lista de figuras

Capítulo 1.- Introducción

1.1 Antecedentes	11
1.2 Objetivo	11
1.3 Producto final	12
1.4 Contribución esperada	12

Capítulo 2.- Bases de datos

2.1 Introducción	13
2.2 ¿Por qué utilizar bases de datos?	14
2.3 Independencia de datos	15
2.4 Arquitectura de un sistema de bases de datos	16

Capítulo 3.- *Data warehouse*

3.1 Concepto	20
3.2 Características	21
3.3 Arquitectura	22
3.4 <i>Data warehouse & Data Mart</i>	25
3.5 Beneficios	26
3.6 Fases de implantación de un <i>data warehouse</i>	26
3.6.1 Definición de los objetivos	27
3.6.2 Definición de los requerimientos de información	28
3.6.3 Diseño y modelización	28
3.6.4 Implementación	28
3.6.5 Revisión	29
3.6.6 Diseño de la estructura de cursos de formación	29
3.7 Ciclo de desarrollo	29
3.7.1 Planeación	30
3.7.2 Requerimientos	30
3.7.3 Análisis	31
3.7.4 Construcción	31
3.7.5 Despliegue	31
3.8 Estrategias de implantación	31
3.9 Factores relevantes en el proceso de adquisición	32
3.9.1 Pruebas en condiciones reales	32
3.9.2 Volumen y organización de los datos	33

3.10 Técnicas de explotación de la implantación	33
3.10.1 <i>Olap, Rolap y Molap</i>	34
3.10.2 <i>Data mining</i> o minería de datos	38
3.10.2.1 Técnicas de <i>Data Mining</i>	38
3.10.2.2 Metodología de aplicación	40
3.10.3 <i>Webhousing</i>	42
3.11 Seguridad de acceso y manipulación de la información en el <i>data warehouse</i>	44
Capítulo 4.- Casos prácticos de <i>data warehouse</i>	
4.1 Caso Office Depot	46
4.2 Caso Delta Air Lines	48
Capítulo 5.- Metodología de la investigación	
5.1 Metodología	50
5.2 Herramienta de investigación (Encuesta)	51
5.3 Variables de la investigación	51
5.4 Tamaño de la muestra	52
Capítulo 6.- Resultados obtenidos y análisis	54
6.1 Información general	55
6.2 Nivel de implementación	56
6.3 Motivos para la implementación	61
6.4 Factores críticos de éxito para la implementación	67
6.5 Problemas más comunes	70
Capítulo 7.- Conclusiones, recomendaciones y trabajos futuros	
7.1 Conclusiones	74
7.1.1 Nivel de implementación	75
7.1.2 Factores críticos de éxito	75
7.1.3 Principales problemas	75
7.1.4 Motivos para la implementación	76
7.2 Recomendaciones	76
7.3 Trabajos futuros	77
Anexo 1	
Formato de la encuesta aplicada	78
Referencias utilizadas	84
Vita	87

Lista de figuras

Fig. 2.1 Composición de una base de datos	13
Fig. 2.2 Los tres niveles de la arquitectura.	17
Fig. 3.1 Componentes del <i>data warehouse</i> y su relación.	20
Fig. 3.2. Arquitectura de un <i>data warehouse</i>	24
Fig. 3.3 Caso de aproximación a una arquitectura descentralizada de <i>data mart</i> .	25
Fig. 3.6 Rapid Warehousing Methodology	27
Fig. 3.7 Ciclo clásico de desarrollo de software	30
Fig. 3.10.1 Explotación de <i>data warehouse</i>	34
Fig. 3.10.2 <i>Web Housing</i>	44
Fig. 6.1 Distribución de las empresas por su giro	54
Fig. 6.2 Número de empleados	55
Fig. 6.3 Ventas anuales en millones de USD	55
Fig. 6.4 TI con que cuentan	55
Fig. 6.5 Nivel que apoyan las TI	55
Fig. 6.6 Empresas que cuentan con <i>data warehouse</i>	56
Fig. 6.7 Área que propuso la implementación del <i>data warehouse</i>	56
Fig. 6.8 Fase de implantación en que se encuentra la herramienta	57
Fig. 6.9 Origen del <i>software</i> de <i>data warehouse</i>	58
Fig. 6.10 Áreas contempladas por el <i>data warehouse</i>	58
Fig. 6.11 Áreas más beneficiadas con el <i>data warehouse</i>	59
Fig. 6.12 Herramientas de explotación usadas	60
Fig. 6.13 Tiempo que duró la implementación	61
Fig. 6.14 Elementos o variables que son necesarios identificar en una empresa para considerar la implementación de un <i>data warehouse</i>	62
Fig. 6.15 Calificación brindada a la productividad operativa generada por el <i>data warehouse</i>	63
Fig. 6.16 Apreciación de la relación costo-beneficio al implementar <i>data warehouse</i>	63
Fig. 6.17 Estudios de factibilidad realizados para la adquisición del <i>data warehouse</i>	64
Fig. 6.18 Tiempo estimado para el retorno de inversión de la implementación de	65

un <i>data warehouse</i>	
Fig. 6.19 Factores de decisión para adquirir la herramienta actual	65
Fig. 6.20 Principales motivos para la implementación de una herramienta de <i>data warehouse</i>	66
Fig. 6.21 Variables o elementos claves para realiza una implementación exitosa	67
Fig. 6.22 Motivos de éxito de <i>data warehouse</i>	68
Fig. 6.23 Involucrados en el éxito de la herramienta	69
Fig. 6.24 Monto aproximado de la inversión en <i>data warehouse</i> en miles de dólares	69
Fig. 6.25 Problemas que enfrentaron el implementar la herramienta	71
Fig. 6.26 Problemas más importantes o que más obstaculizan la implementación de un <i>data warehouse</i>	72



Capítulo 1.- Introducción

1.1 Antecedentes

Tomar mejores decisiones de negocios rápidamente es la llave para el éxito en el mercado competitivo que se vive actualmente. Comprendiendo esto, las organizaciones están buscando mejorar sus sistemas de toma de decisiones ya que pueden ser rebasados por el volumen y la complejidad de los datos disponibles provenientes de sus sistemas de producción y transaccionales. El permitir que todos estos datos estén disponibles para toda la audiencia a lo largo de la empresa es hoy por hoy uno de los retos más significativos para los profesionales involucrados en las tecnologías de información. (Microsoft 1998)

A lo largo del tiempo en el desarrollo de sistemas, se ha tratado de dar solución a esa necesidad intrínseca de los altos mandos de poseer la información de manera resumida o digerida, que le permita en un vistazo corto y rápido, darse cuenta de todo lo que pasa en los sistemas transaccionales de los niveles bajos de la empresa, sin embargo, esto ya no es suficiente, ya que actualmente se requiere conocer lo que sucede en todos y cada uno de los procesos tanto críticos como triviales que se dan lugar en una empresa, ya que es el conocimiento de éstos lo que da una visión clara de lo que ocurre en ella.

Buscando la administración de ese conocimiento tan importante en nuestros días, surge una de las opciones que se han implementado para cubrir esta necesidad de las empresas, que es lo que se conoce desde hace tiempo como *Data Warehouse* (DWH), el cual, es el proceso de extraer datos de diferentes sistemas, funciones, y lugares de la empresa, agregando, filtrando, resumiendo, estandarizando, transformando y clarificando los datos extraídos e integrarlos después en una base consolidada (Anthem Inc, 2001).

1.2 Objetivo

- El objetivo de esta investigación es el de brindar información que permita al lector tener una idea del grado de avance que tienen las empresas grandes en la ciudad de Monterrey, con respecto a la implementación de un sistema de *data warehouse*.
- Identificar cuáles son los motivos principales que hacen pensar en la implementación de un sistema de este tipo.
- Identificar que problemas se presentan en la implementación del *data warehouse* en estas empresas.



- Identificar factores de éxito en la implementación del *data warehouse* en estas empresas.

1.3 Producto final

El producto final que se pretende obtener como resultado de esta investigación, es un documento que presente de manera clara y sencilla algunos aspectos relacionados con la implementación de una herramienta de *data warehouse*.

Así, el lector podrá darse una idea rápida del grado de implementación de herramientas de *data warehouse* que existe en Monterrey, los motivos por los que las empresas consideran que debe implementarse esta herramienta, los factores críticos a considerar en la implementación y por último, los principales problemas que enfrentaron estas empresas en la implementación, lo que podría ser de gran ayuda para cualquier persona que tenga la intención de implementar un *data warehouse*.

1.4 Contribución esperada

Mediante esta investigación, se pretende ofrecer una muestra lo que una herramienta de *data warehouse* representa para las empresas regiomontanas y algunos aspectos a considerarse en su implementación.



Capítulo 2.- Bases de datos

Un ejemplo de una parte potencialmente incomprensible de la terminología de la computadora es el término *base de datos*. Es una palabra inventada, y las dos palabras que la conforman no tienen definiciones absolutamente claras en los contextos cotidianos. Esto se debe en parte a que su significado no es constante: varía considerablemente de acuerdo con el tamaño de la computadora que se utiliza. Por tanto, aunque los usuarios de las microcomputadoras conocen el significado de una base de datos (o lo que por lo general esperan de una), no es el mismo que los usuarios de las macrocomputadoras dan por hecho. (Deakin, 1998)

2.1 Concepto

Una base de datos es un archivo de computadora que tiene una estructura regular formada por registros, que a su vez están formados por campos. En los sistemas de archivo de papel, análogos a los de computadora, cada tarjeta corresponde a un registro, y los campos son los elementos de información como nombre, dirección, fecha, etc. que contienen las tarjetas. Todas las tarjetas de un archivo tienen la misma estructura, con diferentes datos. Las fichas bibliográficas de una biblioteca pueden existir como tarjetas de papel, una por cada libro, o electrónicamente como registros de una base de datos.



Fig. 2.1 Composición de una base de datos

En las bases de datos se manejan normalmente tres tipos de datos: números, caracteres alfabéticos y fechas.

El directorio telefónico, por ejemplo, es una base de datos con tantos registros como números telefónicos existen, con tres campos en cada registro: nombre, dirección y número de teléfono. (ITESO, 1998)

Según Fortier (1997), una base de datos consiste en 4 elementos: datos, relaciones, limitaciones y un esquema. Los datos son una representación computacional binaria de entidades lógicas almacenadas, las relaciones representan una correspondencia entre esos datos, las limitaciones son predicados que definen correctamente los



estados de la base, y el esquema describe la organización de los datos y las relaciones dentro de la base de datos.

2.2 ¿Por qué utilizar bases de datos?

La respuesta general a esta pregunta, es que un sistema de bases de datos proporciona a la empresa un *control centralizado* de sus datos de operación que constituyen uno de los activos mas valiosos. Esto contrasta de manera aguda con la situación que prevalece actualmente en muchas empresas, donde a menudo cada aplicación tiene sus propios archivos, de modo que los datos de operación se encuentran muy dispersos y, por lo tanto, es probable que sean difíciles de encontrar. (Fortier 1997)

Lo anterior implica que en una empresa que utilice un sistema de bases de datos debe existir una persona específica cuya responsabilidad central sea controlar los datos de la operación. Esta persona es el administrador de bases de datos (DBA).

Algunas de las ventajas de tener un control centralizado de los datos son (Fortier, 1997):

- Puede reducirse la redundancia
- Puede evitarse la inconsistencia (al menos en cierta medida)
- Los datos pueden compartirse
- Pueden hacerse cumplir las normas establecidas
- Puede conservarse la integridad
- Pueden equilibrarse los requerimientos contradictorios

Imaginemos un restaurante grande, con un almacén de tal vez miles de botellas y con muchos cambios en ese almacén; o pensemos en una tienda de licor, también con un almacén grande y un gran movimiento en él. En estas situaciones, las ventajas de un sistema computarizado sobre uno tradicional, con métodos basado en papel, donde se guarda el registro de las operaciones, son mas que obvios.

Algunos de los beneficios de usar bases de datos son (Date, 1990):

- Compacticidad: No es necesario la posible voluminosa cantidad de papeles
- Velocidad: la computadora puede cambiar y corregir datos mucho mas rápido de lo que lo hace un humano.
- Menos manipulación: mucho del tedio de mantener los archivos a mano es eliminado
- Actualización: información actualizada esta disponible en cualquier momento que se requiera.



Como podemos observar, todos los beneficios van encaminados de una u otra manera a ser mas eficientes, lo cual, enfocado de le manera adecuada puede ayudar a la reducción de costos, lo que impacta directamente en la función principal de toda empresa que es maximizar su utilidad con el mínimo de esfuerzo y recursos.

2.3 Independencia de los datos

La independencia de los datos puede entenderse con mayor facilidad si primero se analiza el fenómeno opuesto. La mayoría de las aplicaciones actuales son dependientes de los datos. Esto significa que la manera como se organizan los datos en almacenamiento secundario y la manera en que se accesan dependen de los requerimientos de la aplicación y, *además, que el conocimiento de la organización de los datos y de la técnica de acceso forma parte de la lógica de la aplicación*. La aplicación, entonces, debe saber que el incide existe y conocer la secuencia del archivo (como la define aquél), de modo que la estructura interna de la aplicación se construirá con base en este conocimiento.

Se dice que una aplicación como ésta es *dependiente de los datos* porque es imposible cambiar la estructura de almacenamiento (manera como están físicamente registrados los datos) o la estrategia de acceso (manera como se accesan) sin afectar la aplicación.

Es posible definir la independencia de los datos como la *inmunidad de las aplicaciones a los cambios de la estructura de almacenamiento y de la estrategia de acceso*. (Date, 1986)

La ventaja de la independencia de datos, es que permite al usuario concentrarse en la estructura lógica de los datos e ignorar detalles físicos irrelevantes. Por lo tanto, la independencia de los datos, como en un lenguaje de alto nivel de programación es un factor importante para incrementar la productividad de los usuarios (en efecto, independencia de los datos y lenguajes de alto nivel son dos lados de la misma moneda). De hecho, el acceso directo de los usuarios a las bases de datos no sería posible sin algunas características razonables de la independencia de datos. (Date, 1990)



2.4 Arquitectura de un sistema de bases de datos

Los sistemas administradores de bases de datos están compuestos de 5 elementos (Fortier, 1997):

- El hardware, consiste en elementos de procesamiento, memoria volátil, componentes de almacenamiento secundario, dispositivos de almacenamiento de archivos y sensores de entrada.
- El software puede ser dividido en 3 categorías:
 - software de infraestructura de soporte, incluye el sistema operativo y el software para la comunicación por red.
 - software de bases de datos, incluye los componentes de almacenamiento control de concurrencia, procesamiento de transacciones, interfases para la manipulación de las bases de datos, interfase para la definición de las bases de datos y la interfase para el control de las bases de datos.
 - software de aplicaciones. depende de las necesidades del usuario.
- Los datos, son lo que manejan las bases de datos.
- La gente y programas de aplicación, como usuarios, manipulan los datos almacenados, como administradores de bases de datos examina y mantiene las bases de datos para los usuarios.
- Los procedimientos operacionales, son desarrollados y puestos en práctica para dar soporte adicional a los sistemas de bases de datos.

Los componentes principales de un sistema de bases de datos son (Fortier, 1997):

- **Administrador de catálogos.** Mantiene la información de la información almacenada en las bases de datos.
- **Administrador de integridad.** Ayuda en el mantenimiento de las bases de datos, se encarga de la validación, corrección y aseguración de la información contenida.
- **Administrador de transacciones.** Controla y coordina las transacciones dentro de las bases de datos.
- **Administrador de control de concurrencia.** Coordina las acciones de acceso interactivo a la base de datos por transacciones corriendo concurrentemente.
- **Administrados de candados.** Es diseñado para controlar y bloquear los accesos a las bases de datos.
- **Administrador de recuperación.** Debe asegurarse que las bases de datos estén siempre en un estatus de recuperables consistentemente y de manera correcta.



- **Administrador de seguridad.** Tiene la tarea de limitar el acceso, modificaciones, e intrusiones maliciosas a las bases de datos
- **Administrador de soporte al proceso de búsquedas.** Tiene la función de determinar como contestar a los requerimientos de información por parte de los usuarios de la manera mas óptima.
- **Administrador de comunicaciones.** Tiene el rol de policía de transito entre bases de datos
- **Administrador Log.** Tiene el trabajo de coordinar la transferencia de información de la base de datos activa a los almacenamientos persistentes secundarios para ayudar en la recuperabilidad de las bases de datos y mitigar efectivamente los problemas con el sistema operativo.

Como podemos ver en la figura 2.2, la arquitectura se divide en 3 niveles generales: Interno, conceptual y externo.

1. El nivel interno es el mas cercano al almacenamiento físico, es decir, el que concierne a la manera como los datos se almacenan en realidad;
2. El nivel externo es el mas cercano a los usuarios, éste comprende la manera cómo cada usuario ve los datos;
3. El nivel conceptual es un “nivel de mediación” entre los otros dos.

Si el nivel externo se relaciona con las vistas de los usuarios *individuales*, el nivel conceptual puede considerarse como el que define una vista de la *comunidad* de los usuarios. En otras palabras, habrá muchas “vistas externas”, cada una compuesta por una representación mas o menos abstracta de algunas partes de las bases de datos, y habrá una sola “vista conceptual”, compuesta por una representación también abstracta de la base de datos en su totalidad.

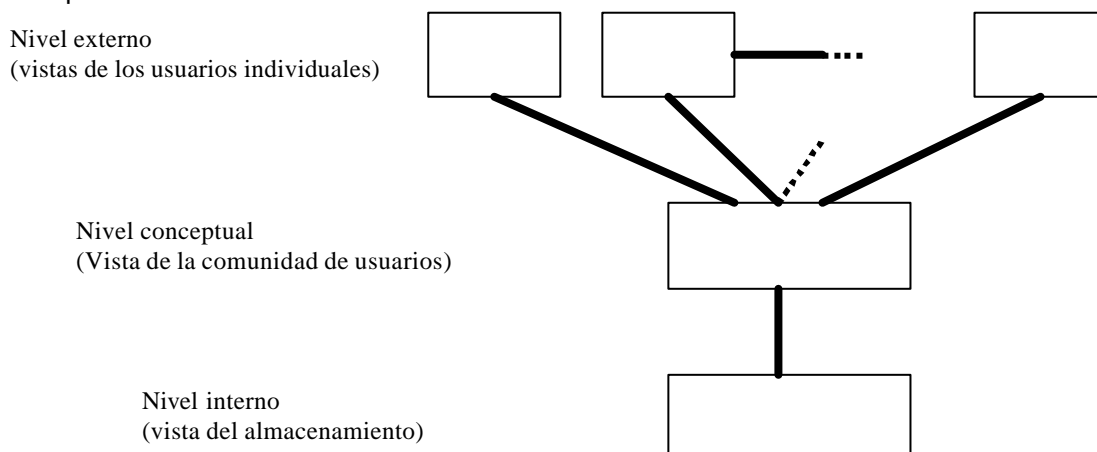


Fig. 2.2 Los tres niveles de la arquitectura.
Fuente: Date. (1986)



Las bases de datos son el componente principal de los *data warehouse*, ya que son los receptáculos de información por excelencia. Existen diferentes maneras de administrar las bases de datos, y diferentes arquitecturas, sin embargo creo que siempre se basará en este tipo de estructuras de almacenamiento por ser las mas sencillas y fáciles de administrar.



Capítulo 3- Data warehouse

Un sistema de *data warehouse*, es simplemente un sencillo, completo y consistente almacén de datos obtenidos de una variedad de fuentes y esta hecho para que los usuarios puedan disponer de él en cualquier momento de manera tal que los datos sean fácilmente entendidos y se conviertan así en una herramienta poderosa en el contexto del negocio.

Lograr la complejidad y consistencia de los datos en el ambiente actual de sistemas de información, es algo bastante complejo. El primer problema es descubrir como definir esa complejidad y consistencia. En el contexto del negocio, esto significa entender las estrategias del negocio y los datos requeridos para apoyar y rastrear sus logros. Este proceso – llamado modelación del negocio – requiere un involucramiento sustancial por parte de los usuarios del negocio y es tradicionalmente un proceso tardado. En los proyectos de *data warehouse*, la modelación del negocio es una fuente vital para el diseño y debe progresar en paralelo con la implementación actual. (Devlin, 2000)

Saber que datos se requieren es el primer paso. Estos datos existen actualmente en varias fuentes de diferentes plataformas, y deben ser copiados de estas fuentes para usarse en el *data warehouse*. Deben ser coordinados de acuerdo al modelo de la empresa, aun cuando estos no hayan sido diseñados para soportar dicha integración. Este debe ser limpiado de errores de estructura y contenido. Este paso – conocido como **población del data warehouse** - es reconocido como uno de los aspectos técnicos mas difíciles en la implementación del *data warehouse*.

Los datos debidamente copiados y transformados de acuerdo al modelo de la empresa son almacenados en el *data warehouse*. Para poder ser entendidos y usados en el contexto del negocio, estos datos deben ser transformados en información. El análisis necesario para hacer esto, ya ha sido desarrollado en el paso de modelación anterior. Lo que los usuarios finales necesitan, es un **catálogo** que describa los datos en su contexto del negocio y que actúe como guía para la localización y uso de esta información.

Finalmente el usuario requiere un conjunto de herramientas para analizar y manipular la información para hacerla útil y disponible. Estas herramientas proveen la interfase ente el usuario y la información, y son el último paso en el proceso de transformación de datos crudos a información útil y usable. (Devlin, 2000)

La figura 3.1 explica los componentes y su relación del *data warehouse*.

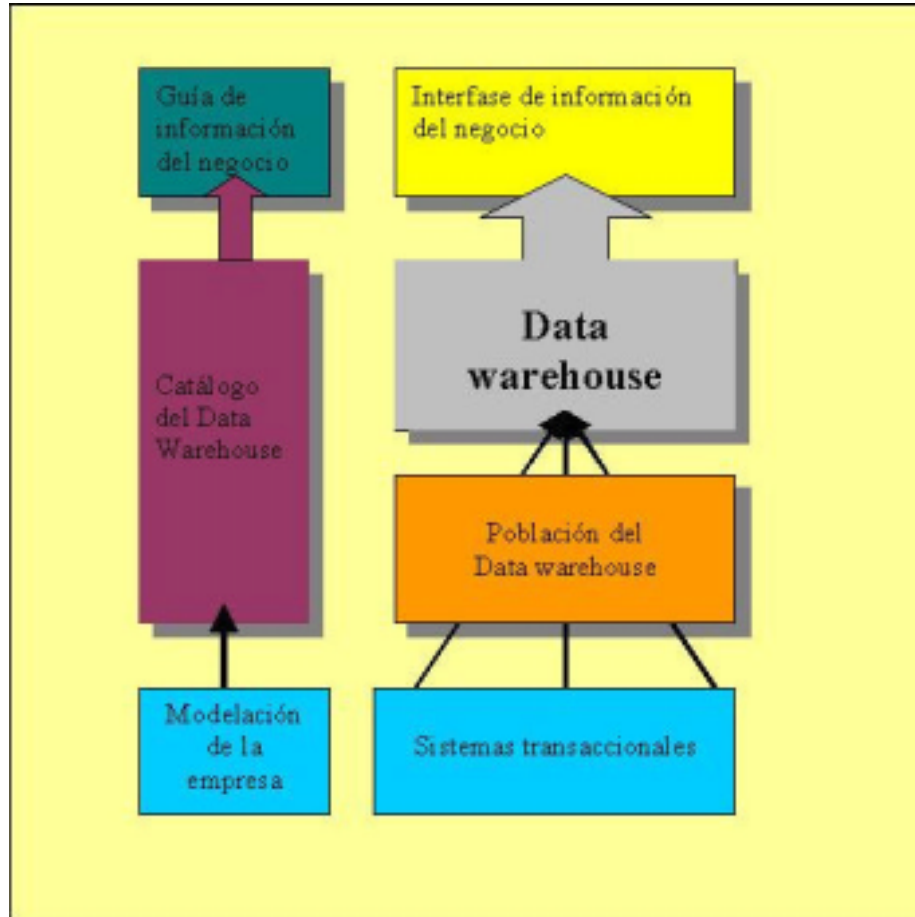


Fig. 3.1 Componentes del *data warehouse* y su relación.
Fuente: Devlin. (2000)

3.1 Concepto

Un *data warehouse* o almacén de datos, puede ser definido de muchas maneras, ya sea como

- una colección de piezas clave de información, usadas para manejar y dirigir los negocios para obtener la mayor rentabilidad (Anahory, Sam y Murray, 1997);
- como conjunto integrado de bases de datos, con orientación temática, que están diseñados para el apoyo a la toma de decisiones, y donde cada unidad de datos es relevante en algún momento del tiempo. (Inmon,1996);
- o como una colección integrada de información corporativa diseñada para la recuperación y el análisis en apoyo a los procesos de toma de decisiones (Kimball, 1996).



En resumen, todos los autores concuerdan que un sistema de *data warehouse* es un almacén de datos referentes a toda la organización, que genera información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales; todo con el fin de apoyar al proceso de toma de decisiones.

Por otro lado, el *data warehousing* o almacenamiento de datos, es el proceso que facilita la creación y explotación de un *data warehouse*, haciendo real la gestión del conocimiento. Para conseguirlo es necesaria la aplicación de una metodología y la implantación de una arquitectura tecnológica.

3.2 Características

Según Inmon (1996), reconocido como el padre del *data warehousing*, el *data warehouse* se caracteriza por ser:

Integrado. Los datos almacenados en el *data warehouse* deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

Temática. Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del *data warehouse*. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

Histórico. El tiempo es parte implícita de la información contenida en un *data warehouse*. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el *data warehouse* sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

No volátil: El almacén de información de un *data warehouse* existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del *data warehouse* la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.



Como podemos ver, la arquitectura del *data warehouse* es algo compleja, pero sirve para garantizar que la información contenida en él, será de utilidad para los usuarios y podrá ser fiable y fácilmente accesada.

3.3 Arquitectura

El *data warehouse*, por ser una colección de datos integrada y orientada a temas, reúne información común de varias fuentes. Como una colección no volátil, variante con el tiempo, el *data warehouse* provee una estable fuente de información en un periodo de tiempo determinado. Un analista de negocios puede usar el *data warehouse* para descubrir que información histórica o actual esta disponible sobre algún tema en particular y entonces obtener la información en particular que necesita, sin tener la limitación de un sistema individual que fue creado para mejorar la eficiencia o efectividad de un proceso del negocio. (Inmon, 1997)

Este tipo de presentación de la información es gracias a las diferentes partes que integran un *data warehouse*, la arquitectura de un *data warehouse* se constituye de un número de partes interconectadas:

Base de datos operacional / Nivel de base de datos externo. Procesan datos para apoyar las necesidades operacionales críticas. Para hacer eso, se han creado las bases de datos operacionales históricas que proveen una estructura de procesamiento eficiente, para un número relativamente pequeño de transacciones comerciales bien definidas. Sin embargo, a causa del enfoque limitado de los sistemas operacionales, las bases de datos diseñadas para soportar estos sistemas, tienen dificultad al acceder a los datos para otra gestión o propósitos informáticos.

Nivel de acceso a la información. Es el nivel del que el usuario final se encarga directamente, incluye el hardware y software involucrados en mostrar información en pantalla y emitir reportes de impresión, hojas de cálculo, gráficos y diagramas para el análisis y presentación.

Nivel de acceso a los datos. Involucrado con el nivel de acceso a la información para conversar en el nivel operacional. El lenguaje de datos común que ha surgido es SQL (Structured Query). Originalmente, SQL fue desarrollado por IBM como un lenguaje de consulta, pero en los últimos veinte años ha llegado a ser el estándar para el intercambio de datos.

Uno de los adelantos claves de los últimos años ha sido el desarrollo de una serie de "filtros" de acceso a datos, tales como EDA/SQL (Enterprise Data Access/ Structured Query Lenguaje) para acceder a casi todos los Sistemas de Gestión de Base de Datos (Data Base Management Systems - DBMSs) y sistemas de archivos de datos, sin



importar si son relacionales o no. Este nivel de acceso a los datos no solamente conecta diferentes DBMSs y sistemas de archivos sobre el mismo hardware, sino también a los fabricantes y protocolos de red, haciéndose responsable de la interfase entre las herramientas de acceso a la información y las bases de datos operacionales.

Nivel de Directorio de Datos (Metadata). Los metadatos es la información o especificaciones alrededor de los datos dentro de la empresa (o datos acerca de los datos) y representa una guía para encontrar las cosas en el *data warehouse*, puesto que definen los formatos, significado y origen de los datos. Un ejemplo sería las sentencias a crear en SQL o un atributo que tiene los siguientes datos: identificador, descripción, tipo de la base de datos, tipo dimensional, formato de despliegue, etc..

A fin de tener un depósito totalmente funcional, es necesario tener una variedad de metadatos disponibles, información sobre las vistas de datos de los usuarios finales e información sobre las bases de datos operacionales.

Nivel de Gestión de Procesos. Tiene que ver con la programación de diversas tareas que deben realizarse para construir y mantener el *data warehouse* y la información del directorio de datos. Este nivel puede depender del control de trabajo para muchos procesos (procedimientos) que deben ocurrir para mantener el *data warehouse* actualizado.

Nivel de Mensaje de la Aplicación. Tiene que ver con el transporte de información alrededor de la red de la empresa. El mensaje de aplicación se refiere también como "subproducto", pero puede involucrar sólo protocolos de red. Puede usarse por ejemplo, para aislar aplicaciones operacionales o estratégicas a partir del formato de datos exacto, recolectar transacciones o los mensajes y entregarlos a una ubicación segura en un tiempo seguro.

Nivel Data Warehouse (Físico). El *data warehouse* (núcleo) es usado principalmente para usos estratégicos. En algunos casos, se puede pensar simplemente como una vista lógica o virtual de datos. En muchos ejemplos, el *data warehouse* puede no involucrar almacenamiento de datos.

Nivel de Organización de Datos. El componente final de la arquitectura *data warehouse* es la organización de los datos. Se llama también gestión de copia o réplica, pero de hecho, incluye todos los procesos necesarios como seleccionar, editar, resumir, combinar y cargar datos en el depósito y acceder a la información desde bases de datos operacionales y/o externas. La organización de datos involucra con frecuencia una programación compleja, pero cada vez más, están creándose las herramientas *data warehousing* para ayudar en este proceso. Involucra también programas de análisis de calidad de datos y filtros que identifican modelos y estructura de datos dentro de los datos operacionales existente.

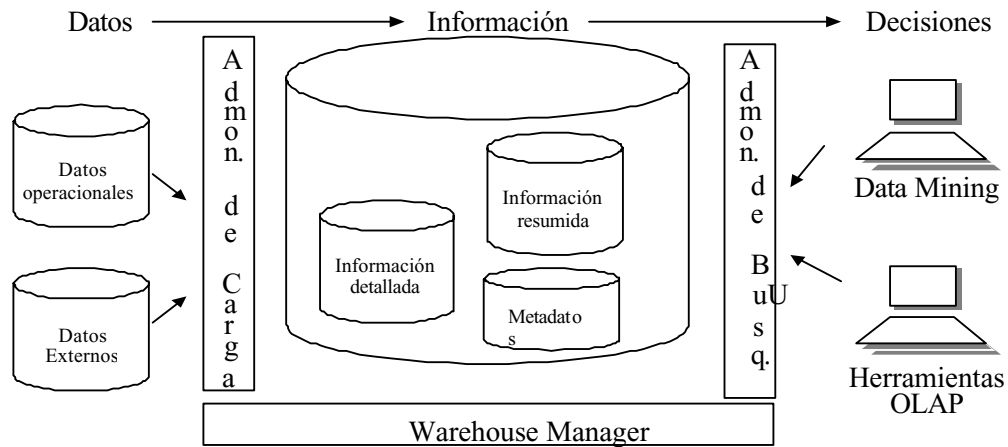


Fig. 3.2. Arquitectura de un *data warehouse*
Fuente: Anahory, Sam y Murray. (1997)

Explicando la figura 3.2 podemos ver que:

Administrador de carga. Es el componente del sistema que ejecuta todas las operaciones necesarias para soportar el proceso de extracción de datos de las diferentes aplicaciones, subirlos a un almacén temporal y carga de los mismos en el *data warehouse* una vez que, previa transformación, cuentan con una estructura similar (construido con diversas herramientas y su tamaño varía dependiendo de la solución de *data warehouse*).

Administrador del Data Warehouse. Se encarga de llevar a cabo todas las operaciones necesarias para la administración del *data warehouse*. Entre sus principales tareas están: verificar la consistencia e integridad de los datos, crear índices y vistas del negocio, generar nuevas agregaciones y actualizar las existentes y respaldar de forma total o parcial los datos del *data warehouse*.

Administrador de búsquedas. Este componente se encarga de llevar a cabo las operaciones para el proceso de administración de consultas (*queries*). Estas son: Dirigir las consultas a las tablas que les correspondan y programar la ejecución de las mismas.

La arquitectura del *data warehouse* está compuesta principalmente de 3 administradores, los cuales se encargan de funciones muy específicas, y que son complementarias para garantizar el debido manejo y funcionamiento del mismo. Las 3 son necesarias y representan la base de toda la arquitectura.



3.4 Data warehouse y data mart

Según define Gray (1998), "un *data mart* es un pequeño *data warehouse* diseñado para una unidad estratégica del negocio (SBU), o un nivel departamental". Los *data marts*, tienen las mismas características de integración, no volatilidad, orientación temática y no volatilidad que el *data warehouse*. Representan una estrategia de "divide y vencerás" para ámbitos muy genéricos de un *data warehouse*.

La función principal de un *data mart* es apoyar a otros sistemas para la toma de decisiones, además de disponer de la información solicitada en el momento preciso (Groth, 2000), permitiendo un mejor control de aquella que se abarca. Esta estrategia es particularmente apropiada cuando el *data warehouse* central crece muy rápidamente y los distintos departamentos requieren sólo una pequeña porción de los datos contenidos en él. La creación de estos *data marts* requiere algo más que una simple réplica de los datos: se necesitarán tanto la segmentación como algunos métodos adicionales de consolidación.

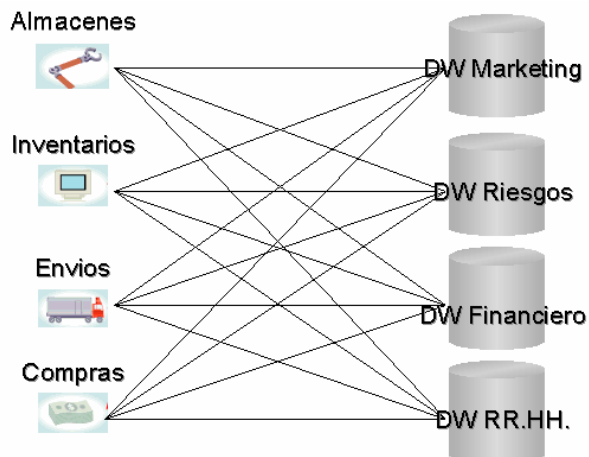


Fig. 3.3 Caso de aproximación a una arquitectura descentralizada de *data mart*.
Fuente: Gray (1998)

El departamento de marketing, emprende el primer proyecto de *data warehouse* como una solución departamental, creando el primer *data mart* de la empresa. Visto el éxito del proyecto, otros departamentos, como el de riesgos, o el financiero se lanzan a crear sus *data marts*. Marketing, comienza a usar otros datos que también usan los *data marts* de riesgos y financiero, y estos hacen lo propio.



Según Gray (1998), en algunos casos, la arquitectura de un *data mart* es muy aconsejable por que:

- Menores cantidades de datos implican que se procesan antes tanto las cargas de datos como las consultas.
- Las peticiones pueden acotarse al área o red que sirve esos datos, sin afectar al resto de usuarios.
- La aplicación cliente que pide la consulta es independiente del servidor que la procesa y del servidor de base de datos que almacena la información

La gran diferencia entre un *data warehouse* y un *data mart* es que éste último está diseñado para un uso específico y no recibe toda la información de la empresa, por lo tanto no tiene una visión completa de la misma.

3.5 Beneficios

Gray (1998) menciona que los beneficios que un *data warehouse* puede aportar son:

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

Es un hecho que el contar con un *data warehouse* es un punto muy importante para cualquier empresa que necesite de un análisis de información muy estructurado y complejo, ya que abarca toda la información de la organización, lo que, entre otras cosas, facilita la aplicación de técnicas de estadística para encontrar relaciones entre productos lo cual puede ser un gran benéfico y reflejarse cuantiosamente en el estado de resultados.

3.6 Fases de implantación de un data warehouse

"Un *data warehouse* no se puede comprar, se tiene que construir". La construcción e implantación de un *data warehouse* es un proceso evolutivo (ComputerWorld 1999). Este proceso se tiene que apoyar en una metodología específica para este tipo de



procesos, el realizar un control para asegurar el seguimiento de la metodología, es más importante que la elección de la mejor de la misma.

En las fases que se establezcan en el alcance del proyecto es fundamental el incluir una fase de formación en la herramienta utilizada para un máximo aprovechamiento de la aplicación. El seguir los pasos de la metodología y el comenzar el *data warehouse* por un área específica de la empresa, nos permitirá obtener resultados tangibles en un corto espacio de tiempo.

La figura 3.6 muestra una metodología propuesta por SAS Institute (1998): la "Rapid Warehousing Methodology". Dicha metodología es iterativa, y está basada en el desarrollo incremental del proyecto de *data warehouse* que se divide en cinco fases.

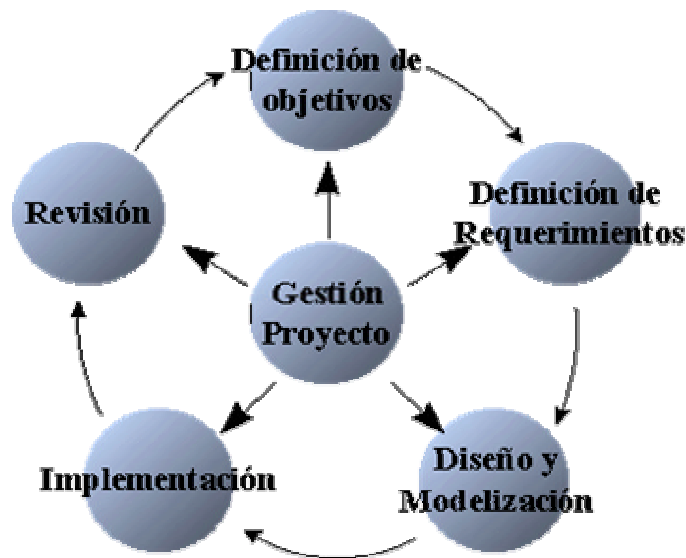


Fig. 3.6 Rapid Warehousing Metodology
Fuente: SAS Institute (1998)

3.6.1-Definición de los objetivos

Es de suma importancia que los objetivos que se buscan con la implementación de un *data warehouse* estén bien definidos, de lo contrario podría convertirse en un proyecto inmanejable, muy costoso y contraproducente.

Los objetivos deben ser reales, alcanzables y medibles, y deben estar bien alineados con la estrategia del negocio para que pueda ser redituable la inversión en un proyecto de este tamaño.



3.6.2-Definición de los requerimientos de información

Tal como sucede en todo tipo de proyectos, sobre todo si involucran técnicas novedosas como son las relativas al *data warehouse*, lo primero es analizar las necesidades y hacer comprender las ventajas que este sistema puede reportar.

Esta es la etapa en donde el usuario va a jugar un papel muy destacado, ya que será quien realice las consultas y explote propiamente la información que podrá obtenerse del *data warehouse*, es de suma importancia que comprenda claramente los beneficios que traerá la implementación de esta herramienta.

3.6.3-Diseño y modelización

Los requerimientos de información identificados durante la anterior fase proporcionarán las bases para realizar el diseño y la modelización del *data warehouse*.

En esta fase se identificarán las fuentes de los datos (sistema operacional, fuentes externas,..) y las transformaciones necesarias para que, a partir de dichas fuentes, se obtenga el modelo lógico de datos del *data warehouse*. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades de negocio de la organización.

El modelo lógico se traducirá posteriormente en el modelo físico de datos que se almacenará en el *data warehouse* y que definirá la arquitectura de almacenamiento del *data warehouse* adaptándose al tipo de explotación que se realice del mismo.

La mayor parte estas definiciones de los datos del *data warehouse* estarán almacenadas en los metadatos y formarán parte del mismo.

3.6.4-Implementación

La implantación de un *data warehouse* lleva implícitos los siguientes pasos:

- Extracción de los datos del sistema operacional y transformación de los mismos.
- Carga de los datos validados en el *data warehouse*. Esta carga deberá ser planificada con una periodicidad que se adaptará a las necesidades de refresco detectadas durante las fases de diseño del nuevo sistema.
- Explotación del *data warehouse* mediante diversas técnicas dependiendo del tipo de aplicación que se de a los datos:



- ✓ Consultas y Reportes
- ✓ Procesos analíticos en línea (OLAP)
- ✓ Sistema de Información Ejecutiva (EIS)
- ✓ Sistema de Apoyo a las Decisiones (DSS)
- ✓ Visualización de la información
- ✓ Minería de Datos, etc.

La información necesaria para mantener el control sobre los datos se almacena en los metadatos técnicos (cuando describen las características físicas de los datos) y de negocio (cuando describen cómo se usan esos datos). Dichos metadatos deberán ser accesibles por los usuarios finales que permitirán en todo momento tanto al usuario, como al administrador que deberá además tener la facultad de modificarlos según varíen las necesidades de información.

Con la finalización de esta fase se obtendrá un *data warehouse* disponible para su uso por parte de los usuarios finales y el departamento de informática.

3.6.5-Revisión

La construcción del *data warehouse* no finaliza con la implantación del mismo, sino que es una tarea iterativa en la que se trata de incrementar su alcance aprendiendo de las experiencias anteriores.

Después de implantarse, debería realizarse una revisión del *data warehouse* planteando preguntas que permitan, después de los seis o nueve meses posteriores a su puesta en marcha, definir cuáles serían los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.

3.6.6-Diseño de la estructura de cursos de formación

Con la información obtenida de reuniones con los distintos usuarios se diseñarán una serie de cursos a medida, que tendrán como objetivo el proporcionar la formación estadística necesaria para el mejor aprovechamiento de la funcionalidad incluida en la aplicación. Se realizarán prácticas sobre el desarrollo realizado, las cuales permitirán fijar los conceptos adquiridos y servirán como formación a los usuarios.

3.7 Ciclo de desarrollo

El *data warehouse* sigue el mismo ciclo de perfeccionamiento que todos los desarrollos de software. Las fases del ciclo son las mismas, lo mismo que su secuencia, sólo existen variantes únicas que se relacionan específicamente con el *data*



warehouse para tareas dentro de estas fases. La figura 3.7 muestra el ciclo clásico de desarrollo de software:

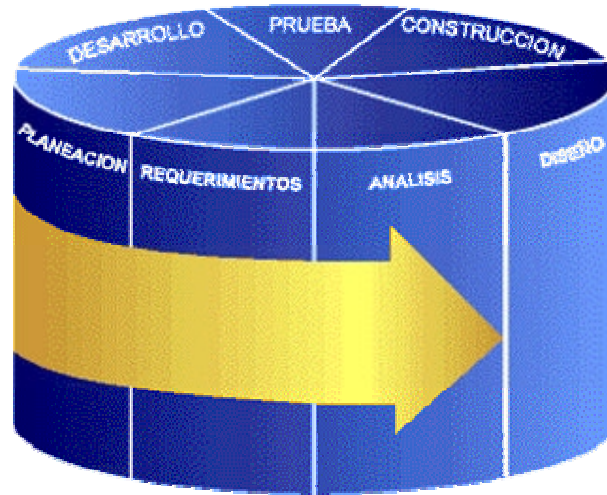


Fig. 3.7 Ciclo clásico de desarrollo de software
Fuente: Inmon. (1996)

3.7.1 Planeación

La planeación es una fase importante de la implementación del *data warehouse*. Las decisiones tomadas durante la fase de planeación tienen un impacto significativo en el ámbito de implementación y en la magnitud del esfuerzo. Las decisiones clave de planeación incluyen la selección de un enfoque de arriba hacia abajo (de lo general a lo particular), de abajo hacia arriba (en sentido opuesto) o combinado; la selección de la arquitectura apropiada del *data warehouse*; la selección adecuada del ámbito de información, fuentes de datos y tamaño del metamodelo; y la estimación de planes de programas y proyectos y justificaciones de presupuesto. (Kimball, 1998)

3.7.2 Requerimientos

Durante la fase de requerimientos se debe considerar una diversidad de ellos. Los requerimientos son conducidos por el negocio y por la tecnología. La cuidadosa selección y especificación de requerimientos en esta etapa proporciona un proyecto cimentado que arroja resultados con rapidez. (Inmon 1996)



3.7.3 Análisis

La fase de análisis es importante ya que determina la forma en que se cubrirán los requerimientos. Esta fase se enfoca principalmente en la conversión de especificaciones de requerimientos a especificaciones de metamodelo para el *data warehouse*. Después, estas especificaciones se usan para generar extractores del *data warehouse* y software de transformación, integración, resumen y adición. (Kimball, 1996)

3.7.4 Construcción

La fase de construcción resalta los diversos intercambios "construir en comparación con comprar". Mediante la selección adecuada de componentes suministrados por fabricantes, es posible construir una primera implementación del *data warehouse* rápida y eficaz. (Inmon 1996)

3.7.5 Despliegue

La fase de despliegue en el ciclo de desarrollo del *data warehouse* tiene un componente único denominado comercialización de información. Esto reconoce que la mercancía que suministra el *data warehouse* a sus usuarios finales (clientes) es la propia información. Como un producto de mercancía, la información también debe comercializarse como los bienes de consumo. La comercialización comprende la capacidad de hacer énfasis en la disponibilidad, los beneficios y el empaque para hacerla atractiva al usuario final. (Kimball, 1998)

3.8 Estrategias de implantación

Consideraciones que deben seguirse a la hora de abordar un proyecto de este tipo (Debevoise 1999):

"La Base de Datos de Riesgos debe estar separada de las Bases de Datos Operacionales" con objeto de no interferir en la actividad del día a día, disponiendo de la información necesaria para Riesgos (interna y externa) y en un entorno orientado hacia la consulta y el análisis (*data warehouse*).

"Concepción del sistema como un conjunto de herramientas de análisis", debido a que las actividades de Análisis de Riesgos no se pueden automatizar completamente, puesto que requieren análisis y decisiones del usuario.

"Diseño del sistema no orientado a procesos"; se debe disponer de un conjunto abierto de herramientas que se utilizan con propósitos determinados no relacionados con las necesidades operativas.



"Abordar el sistema con un enfoque de desarrollo gradual", se debe comenzar con un esqueleto básico de funcionalidad y datos que produzcan resultados a corto plazo y permita aprender en la práctica, y a continuación ir configurando progresivamente nuevas funcionalidades conforme la experiencia lo vaya requiriendo.

3.9 Factores relevantes en el proceso de adquisición

En la definición del objeto del contrato y los requisitos inherentes al mismo, así como en la valoración y comparación de ofertas de los licitadores pueden intervenir muchos factores y de muy diversa índole.

Es de suma importancia que todos los factores relevantes que intervienen en el proceso de contratación queden debidamente recogidos en el pliego de prescripciones técnicas que regule el contrato. Así mismo, es conveniente que las soluciones ofertadas por los licitadores sean recogidas en los cuestionarios disponibles a tal efecto:

- De empresa
- Económicos
- Técnicos particulares

Se van a relacionar a continuación algunos de los factores que suelen tener mayor peso al seleccionar una herramienta de *data warehouse*. Sin embargo, debe tenerse en cuenta que la importancia de cada factor variará en función de cada caso particular, por lo que siempre será necesario identificar la importancia relativa de cada punto. (Gray, 1998)

3.9.1 Pruebas en condiciones reales

El rendimiento real de un *data warehouse* es muy difícil de predecir mediante procedimientos teóricos. Si se va a instalar un *data warehouse* que contendrá un gran volumen de datos o, si por cualquier otra razón, existen dudas sobre la capacidad del *data warehouse* de dar unas prestaciones adecuadas en las máquinas disponibles se debe exigir al suministrador una prueba anterior a la adquisición del *data warehouse*. Esta prueba debe realizarse en la propia instalación de destino.

La prueba se debería realizar en las condiciones más parecidas a las reales que se puedan conseguir. Para ello se deberá cargar el *data warehouse* con un volumen de datos adecuado y se deberán crear procesos de prueba similares a los más costosos de los que se vayan a desarrollar. (Kimball, 1996)

No es preciso realizar la prueba en momentos de gran carga, por la diferente filosofía de un almacén de datos orientado al conocimiento, pero sí que será preciso la



comprobación de la compatibilidad de la herramienta para los procesos de extracción y carga desde los diferentes sistemas operacionales (sistemas operativos, bases de datos, etc.) implicados.

3.9.2 Volumen y organización de los datos

Debe estar garantizado que el *data warehouse* es capaz de tratar el volumen de datos que se va a necesitar en la instalación. Para ello debe verificarse no sólo que el *data warehouse* puede manejar el volumen total de datos, sino que no existe ninguna limitación que impide organizarlo de la forma más conveniente.

No obstante, cabe reseñar que muchos problemas de rendimiento se deben más veces a un mal diseño del modelo de datos del *data warehouse* que a un problema de rendimiento de la herramienta en sí. (Inmon, 1999)

El decidir cuando implementar un sistema de *data warehouse* no puede ser una decisión del departamento de sistemas, debe involucrarse a todos los niveles de la organización para garantizar el buen desarrollo e implementación. Es un hecho que el nivel al que hay que convencer no puede ser solamente uno, ya que de no ser aceptado por todos, no importa que tan bueno sea, estará destinado al fracaso. Es muy importante identificar los factores importantes del proceso de adquisición, para saber con certeza lo que se espera del *data warehouse*.

3.10 Técnicas de explotación de la implantación

Existen diferentes herramientas para explotar la información que contiene un *data warehouse*. Dependiendo del tipo de consulta que se desea, es el tipo de herramienta que debe seleccionarse.

A continuación presentaré una breve descripción de las herramientas más comunes para explotar el *data warehouse*.



3.10.1.- Olap. Rolap, Molap

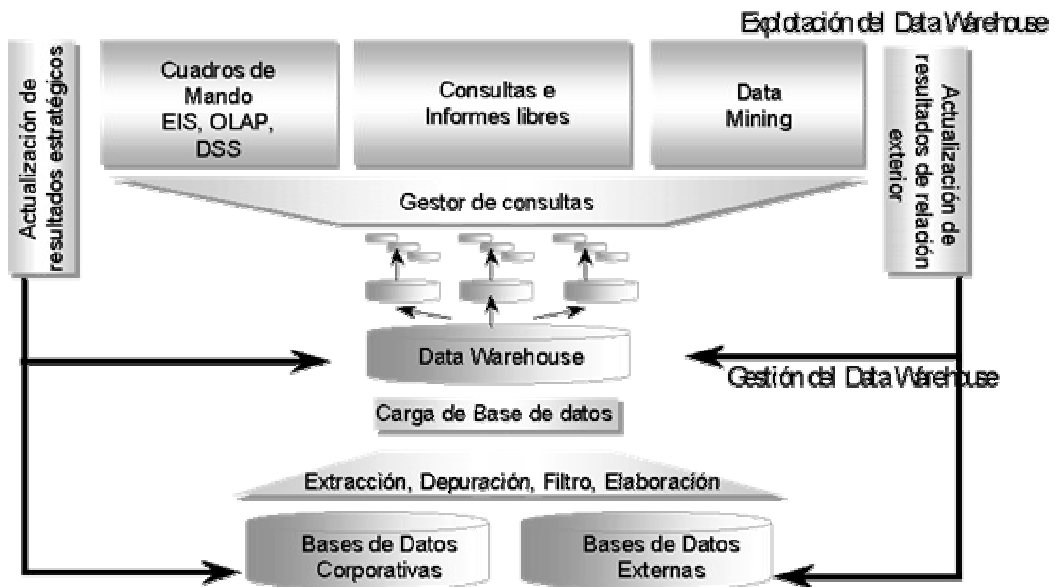


Fig. 3.10.1 Explotación de *data warehouse*
Fuente: Devlin. (2000)

La explotación del *data warehouse* mediante información de gestión, se fundamenta básicamente en los niveles agrupados o calculados de información.

La información de gestión se compone de conceptos de información y coeficientes de gestión, que los cuadros directivos de la empresa pueden consultar según las dimensiones de negocio que se definan.

Dichas dimensiones de negocio se estructuran a su vez en distintos niveles de detalle (por ejemplo, la dimensión geográfica puede constar de los niveles nacional, provincial, ayuntamientos y sección censal).

Este tipo de sistemas ha existido desde hace tiempo, en el mundo de la informática bajo distintas denominaciones: cuadros de mando, Sistemas de información administrativa (MIS), Sistemas de información ejecutiva (EIS), etc.

Su realización fuera del entorno del *data warehouse*, puede repercutir sobre estos sistemas en una mayor rigidez, dificultad de actualización y mantenimiento, malos tiempos de respuesta, incoherencias de la información, falta del dato agregado, etc.



Sistemas OLAP

Los sistemas de soporte a la decisión usando tecnologías de *data warehouse*, se llaman sistemas OLAP (siglas de On Line Analytical Processing (OLAP)). En general, estos sistemas OLAP deben:

- ❑ Soportar requerimientos complejos de análisis
- ❑ Analizar datos desde diferentes perspectivas
- ❑ Soportar análisis complejos contra un volumen grande de datos

La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y algunas posibilidades de navegación, seleccionando la información a obtener. (Fernández, 1999)

Normalmente este tipo de selecciones se ve reflejada en la visualización de la estructura multidimensional, en unos campos de selección que nos permitan elegir el nivel de agregación (jerarquía) de la dimensión, y/o la elección de un dato en concreto, la visualización de los atributos del sujeto, frente a una(s) dimensiones en modo tabla.

Existen dos arquitecturas diferentes para los sistemas OLAP: OLAP multidimensional (MOLAP) y OLAP relacionales (ROLAP).

Sistemas MOLAP

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente. Por el contrario, la arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales. (Anahory, Sam y Murray, 1997)

Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada multidimensionalmente.

El sistema MOLAP utiliza una arquitectura de dos niveles: La bases de datos multidimensionales y el motor analítico.

- La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.
- El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona un interfaz a través del cual los usuarios finales visualizan los análisis OLAP. Una



arquitectura cliente/servidor permite a varios usuarios acceder a la misma base de datos multidimensional.

La información procedente de los sistemas operacionales, se carga en el sistema MOLAP, mediante una serie de rutinas batch. Una vez cargado el dato elemental en la Base de Datos multidimensional (MDDDB), se realizan una serie de cálculos en batch, para calcular los datos agregados, a través de las dimensiones de negocio, rellenando la estructura MDDDB.

Tras rellenar esta estructura, se generan unos índices y algoritmos de tablas hash para mejorar los tiempos de accesos a las consultas.

Una vez que el proceso de compilación se ha acabado, la MDDDB está lista para su uso. Los usuarios solicitan informes a través del interface, y la lógica de aplicación de la MDDDB obtiene el dato.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos precompilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan precalculados y almacenados previamente. (Gray, 1998)

Sistemas ROLAP

La arquitectura ROLAP, accede a los datos almacenados en un *data warehouse* para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales.

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica.

- El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato.
- El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.
- El motor ROLAP se integra con niveles de presentación, a través de los cuales los usuarios realizan los análisis OLAP.

Después de que el modelo de datos para el *data warehouse* se ha definido, los datos se cargan desde el sistema operacional. Se ejecutan rutinas de bases de datos para agregar el dato, si así es requerido por el modelos de datos y se crean entonces los índices para optimizar los tiempos de acceso a las consultas.



Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.

La arquitectura ROLAP es capaz de usar datos precalculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del *data warehouse*, y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y *joins* múltiples. (ATG, 1999)

ROLAP vs. MOLAP (Comparativa)

Cuando se comparan las dos arquitecturas, se pueden realizar las siguientes observaciones (Gray 1998):

El ROLAP delega la negociación entre tiempo de respuesta y el proceso batch al diseño del sistema. Mientras, el MOLAP, suele requerir que sus bases de datos se precompilen para conseguir un rendimiento aceptable en las consultas, incrementando, por tanto los requerimientos batch.

Los sistemas con alta volatilidad de los datos (aquellos en los que cambian las reglas de agregación y consolidación), requieren una arquitectura que pueda realizar esta consolidación ad-hoc. Los sistemas ROLAP soportan bien esta consolidación dinámica, mientras que los MOLAP están más orientados hacia consolidaciones batch.

Los ROLAP pueden crecer hasta un gran número de dimensiones, mientras que los MOLAP generalmente son adecuados para diez o menos dimensiones. Los ROLAP soportan análisis OLAP contra grandes volúmenes de datos elementales, mientras que los MOLAP se comportan razonablemente en volúmenes más reducidos (menos de 5 Gb)

Por ello, y resumiendo, el ROLAP es una arquitectura flexible y general, que crece para dar soporte a amplios requerimientos OLAP. El MOLAP es una solución particular, adecuada para soluciones departamentales con unos volúmenes de información y número de dimensiones más modestos (Gray 1998).



3.10.2 Data mining o minería de datos

El *Data Mining* es un proceso que, a través del descubrimiento y cuantificación de relaciones predictivas en los datos, permite transformar la información disponible en conocimiento útil de negocio. Esto es debido a que no es suficiente "navegar" por los datos para resolver los problemas de negocio, sino que se hace necesario seguir una metodología ordenada que permita obtener rendimientos tangibles de este conjunto de herramientas y técnicas de las que dispone el usuario.

Constituye por tanto una de las vías clave de explotación del *data warehouse*, dado que es este su entorno natural de trabajo. (Groth, 2000)

Se trata de un concepto de explotación de naturaleza radicalmente distinta a la de los sistemas de información de gestión, dado que no se basa en coeficientes de gestión o en información altamente agregada, sino en la información de detalle contenida en el almacén. Adicionalmente, el usuario no se conforma con la mera visualización de datos, sino que trata de obtener una relación entre los mismos que tenga repercusiones en su negocio. (NorSistemas, 1998)

3.10.2.1 Técnicas de data mining

Para soportar el proceso de *data mining*, el usuario dispone de una extensa gama de técnicas que le pueden ayudar en cada una de las fases de dicho proceso:

Para realizar un análisis estadístico, se cuenta con las siguientes herramientas (ATG, 2000):

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

Regresión: define la relación entre una o más variables y un conjunto de variables predictoras de las primeras.

Ji cuadrada: contrasta la hipótesis de independencia entre variables.

Componentes principales: permite reducir el número de variables observadas a un menor número de variables artificiales, conservando la mayor parte de la información sobre la varianza de las variables.

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



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

Ahora que si el usuario prefiere usar métodos basados en árboles de decisión, cuenta con las siguientes herramientas (LGI, 1998):

El método Chaid (Chi Squared Automatic Interaction Detector) es un análisis que genera un árbol de decisión para predecir el comportamiento de una variable, a partir de una o más variables predictoras, de forma que los conjuntos de una misma rama y un mismo nivel son disjuntos. Es útil en aquellas situaciones en las que el objetivo es dividir una población en distintos segmentos basándose en algún criterio de decisión.

El árbol de decisión se construye partiendo el conjunto de datos en dos o más subconjuntos de observaciones a partir de los valores que toman las variables predictoras. Cada uno de estos subconjuntos vuelve después a ser particionado utilizando el mismo algoritmo. Este proceso continúa hasta que no se encuentran diferencias significativas en la influencia de las variables de predicción de uno de estos grupos hacia el valor de la variable de respuesta.

La raíz del árbol es el conjunto de datos íntegro, los subconjuntos y los subconjuntos conforman las ramas del árbol. Un conjunto en el que se hace una partición se llama nodo.

El número de subconjuntos en una partición puede ir de dos hasta el número de valores distintos que puede tomar la variable usada para hacer la separación. La variable de predicción usada para crear una partición es aquella más significativamente relacionada con la variable de respuesta de acuerdo con test de independencia de la Chi cuadrada sobre una tabla de contingencia.

Algoritmos genéticos:

Son métodos numéricos de optimización, en los que aquella variable o variables que se pretenden optimizar junto con las variables de estudio constituyen un segmento de información. Aquellas configuraciones de las variables de análisis que obtengan mejores valores para la variable de respuesta, corresponderán a segmentos con mayor capacidad reproductiva. A través de la reproducción, los mejores segmentos perduran y su proporción crece de generación en generación.

Se puede además introducir elementos aleatorios para la modificación de las variables (mutaciones). Al cabo de cierto número de iteraciones, la población estará constituida por buenas soluciones al problema de optimización.



Redes neuronales:

Genéricamente son métodos de proceso numérico en paralelo, en el que las variables interactúan mediante transformaciones lineales o no lineales, hasta obtener unas salidas. Estas salidas se contrastan con los que tenían que haber salido, basándose en unos datos de prueba, dando lugar a un proceso de retroalimentación mediante el cual la red se reconfigura, hasta obtener un modelo adecuado.

Lógica difusa:

Es una generalización del concepto de estadística. La estadística clásica se basa en la teoría de probabilidades, a su vez ésta en la técnica conjuntista, en la que la relación de pertenencia a un conjunto es dicotómica (el 2 es par o no lo es). Si establecemos la noción de conjunto borroso como aquel en el que la pertenencia tiene una cierta graduación (¿un día a 20°C es caluroso?), dispondremos de una estadística más amplia y con resultados más cercanos al modo de razonamiento humano.

Series temporales

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

Como podemos ver, el data mining tiene una gran variedad de técnicas, de entre las cuales se va a elegir la que mas le convenga al usuario, dependiendo del tipo de información que requiere. Con estas técnicas se convierte la información aislada y sin gran importancia en información útil que puede cambiar el curso de un negocio u organización.

3.10.2.2 Metodología de aplicación:

Para utilizar estas técnicas de forma eficiente y ordenada es preciso aplicar una metodología estructurada, al proceso de *Data Mining*.

Muestreo

Extracción de la población muestral sobre la que se va a aplicar el análisis. En ocasiones se trata de una muestra aleatoria, pero puede ser también un subconjunto de datos del *data warehouse* que cumplan unas condiciones determinadas. El objeto de trabajar con una muestra de la población en lugar de toda ella, es la simplificación



del estudio y la disminución de la carga de proceso. La muestra más óptima será aquella que teniendo un error asumible contenga el número mínimo de observaciones. (Groth, 2000)

En el caso de que se recurra a un muestreo aleatorio, se debería tener la opción de elegir:

- El nivel de confianza de la muestra (usualmente el 95% o el 99%).
- El tamaño máximo de la muestra (número máximo de registros), en cuyo caso el sistema deberá informar del el error cometido y la representatividad de la muestra sobre la población original.
- El error muestral que está dispuesto a cometer, en cuyo caso el sistema informará del número de observaciones que debe contener la muestra y su representatividad sobre la población original.

Para facilitar este paso se debe disponer de herramientas de extracción dinámica de información con o sin muestreo (simple o estratificado). En el caso del muestreo, dichas herramientas deben tener la opción de, dado un nivel de confianza, fijar el tamaño de la muestra y obtener el error o bien fijar el error y obtener el tamaño mínimo de la muestra que nos proporcione este grado de error. (Adiaans, 1997)

Exploración

Una vez determinada la población que sirve para la obtención del modelo se deberá determinar cuales son las variables explicativas que van a servir como "inputs" al modelo. Para ello es importante hacer una exploración por la información disponible de la población que nos permita eliminar variables que no influyen y agrupar aquellas que repercuten en la misma dirección.

El objetivo es simplificar en lo posible el problema con el fin de optimizar la eficiencia del modelo. En este paso se pueden emplear herramientas que nos permitan visualizar de forma gráfica la información utilizando las variables explicativas como dimensiones. (Adiaans, 1997)

También se pueden emplear técnicas estadísticas que nos ayuden a poner de manifiesto relaciones entre variables. A este respecto resultará ideal una herramienta que permita la visualización y el análisis estadístico integrados:

Manipulación

Tratamiento realizado sobre los datos de forma previa a la modelización, en base a la exploración realizada, de forma que se definan claramente los inputs del modelo a realizar (selección de variables explicativas, agrupación de variables similares, etc.). (Groth, 2000)



Modelización

Permite establecer una relación entre las variables explicativas y las variables objeto del estudio, que posibilitan inferir el valor de las mismas con un nivel de confianza determinado.

Valoración

Análisis de la bondad del modelo contrastando con otros métodos estadísticos o con nuevas poblaciones muestrales.

El proceso de *data mining* es muy poderoso y útil, sin embargo, para hacerlo mas eficiente debemos aplicar una metodología de manera adecuada para que las técnicas sean controladas y arrojen los resultados que estamos esperando, de lo contrario la información podría ser inmanejable y por ende inútil.

3.10.3.- WEBHOUSING

La popularización de Internet y la tecnología Web, ha creado un nuevo esquema de información en el cual los clientes tienen a su disposición unas cantidades ingentes de información. La integración de las tecnologías Internet y *data warehouse* tienen una serie de ventajas como son (Mattison, 1999):

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

La interactividad de las aplicaciones en este entorno pueden tener varios niveles (LGI, 1998):

- Publicación de datos: las páginas distribuyen información obtenida del *data warehouse*, volcada en las páginas intra/internet.



- Distribución de reportes: dando soporte a consultas simples elaboradas por los usuarios.
- Aplicaciones dinámicas: sirviendo de soporte de decisión a servicios solicitados desde el puesto cliente, ejecutando la petición en el servidor y devolviéndolas al cliente, vía el *browser* de Internet o haciendo uso de "*applets*" de Java.

Las arquitecturas base de una implantación de *data warehouse* en Internet, pueden tener las siguientes alternativas (ATG, 2000):

1. Usar el Servidor Internet como router, y ejecutar la petición desde el cliente al servidor directamente.
2. Hacer uso del navegador para visualizar una página Internet residente en el servidor de Internet. Esta página contendría información que se actualizaría en el servidor Internet, desde el servidor DW, a petición del usuario haciendo uso de CGI's.
3. El cliente podría lanzar su consulta directamente al servidor de DW, con "*applets*" de Java, haciendo el servidor Internet únicamente de encaminamiento (router).
4. El cliente podría ejecutar la aplicación DW desde el navegador, pero con un *plug-in*, que haría que se tuvieran las mismas opciones que la aplicación DW.
5. Realizar una descarga masiva de datos con un protocolo de transferencia de archivos (FTP), para su proceso en local.

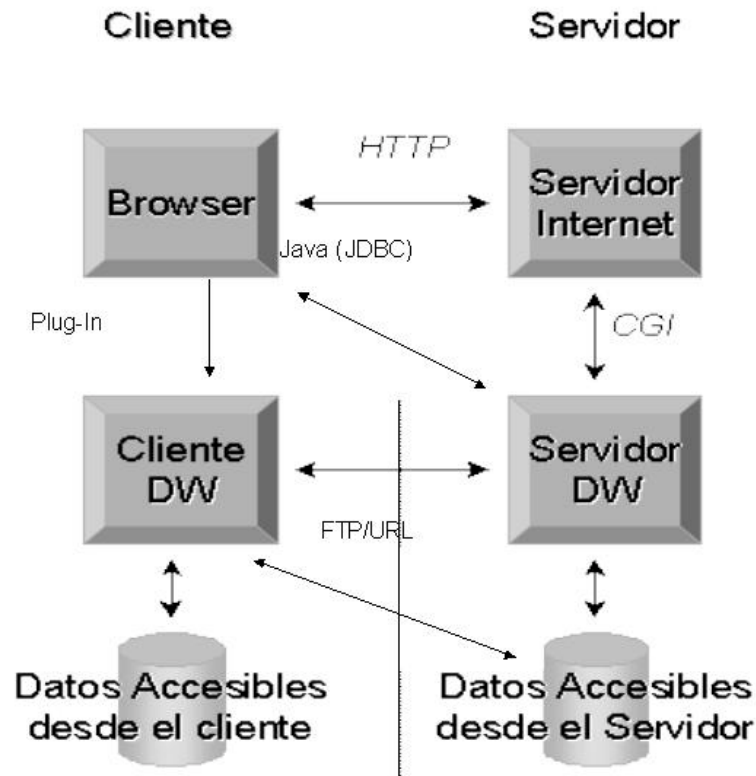


Fig. 3.10.2 Web Housing
Fuente: ATG. (2000)

El alcance funcional de la implantación del *data warehouse*, basado en tecnologías Internet, puede ser la misma que la realizada sin su uso. En este sentido las críticas que se le pueden achacar en la actualidad, provienen de la baja velocidad de las líneas actuales, que se solventa parcialmente mediante el uso de aplicaciones Java, en lugar de hacer uso de páginas HTML (*Hypertext Markup Language*), o CGI (*Common Gateway Interface*). Solución parcial, mientras la velocidad de transferencia se incrementa día a día mediante nuevos algoritmos de compresión de datos o el uso de líneas de alta capacidad RDSI (Red Digital de Servicios Integrados).

3.11. SEGURIDAD DE ACCESO Y MANIPULACIÓN DE LA INFORMACIÓN EN EL DW

Si bien la seguridad de accesos (al nivel de datos y de aplicación) debe ser tratada de la misma manera que en los sistemas operacionales, los procedimientos de copias de seguridad merecen un especial tratamiento.



Tal y como ocurre en los sistemas operacionales, un sistema *data warehouse* debe poder realizar procedimientos de recuperación de la información desde cualquier momento en el que los datos estaban validados. Un *data warehouse*, debe poder contar con procedimientos de recuperación, que permitan recuperar los datos ante cualquier situación de catástrofe.

No obstante, es preciso tener en cuenta otras consideraciones, así por ejemplo dependiendo del tamaño de un Data Mart, se puede elegir no realizar un backup, sino realizar un refresco especial desde los datos operacionales, dependiendo de la periodicidad estándar de carga. (Gray, 1998)

En cuanto a la seguridad de acceso, se cumple en los sistemas de *data warehouse*, que es preciso el implantar niveles de acceso a la información, realizando un plan completo de seguridad que contemple (LGI, 1998):

- Acceso a recursos de la red (local o intranet)
- Asignación de usuarios a grupos con perfiles de seguridad diferenciados
- Asignación de niveles de autorización de aplicación a grupos de usuarios
- Seguridad a nivel de Base de Datos, mediante los procedimientos provistos por las mismas.
- Etc.

Definitivamente la seguridad siempre jugará un papel importante siempre que se hable de datos, lo interesante es saber decidir que tipo de seguridad y recuperación se va a emplear, de manera que se pueda combinar la robustez de un sistema hermético, con el dinamismo y flexibilidad de respaldo de un sistema pequeño.



Capítulo 4.- Casos prácticos

A continuación se muestran algunos casos prácticos que nos darán una idea de lo que una herramienta de *data warehouse* puede hacer por una empresa en caso de ser bien planeado e implementado.

4.1 Caso Office Depot

Como el proveedor mas grande del mundo de productos de oficina a menudeo y negocios de todos los tamaños, Office Depot lidera su industria en todos los canales de distribución. Office Depot actualmente opera mas de 950 tiendas y hace negocios en 18 países alrededor del mundo. Desde su presencia solo de manera presencial (*brick and mortar*), Office Depot ha migrado a múltiples canales, ofreciendo compras a través de internet por medio de su sitio *web*, catálogos de compra y contratación de soporte incluyendo sitios personalizados para clientes corporativos. (NCR, 2000)

Fundada en 1986, Office Depot ha llegado a su posición de liderazgo en la industria de productos de oficina creando la misión de traer el producto correcto, al precio correcto en el tiempo correcto para sus clientes.

Sus ventas han incrementado de \$10.3 billones en 1999 a \$11.6 billones en el 2000, lo que representa un 13% de incremento en ventas.

Con mas de 2,000 camiones, 1,000 manejadores de cuentas, 60 locales para oficinas de ventas, 25 centros de entrega doméstica, 8 *call centers* regionales, y 2 sitios en internet, Office Depot ha alcanzado su misión.

Situación

A medida que crecía su cantidad de clientes, Office Depot se dió cuenta de la necesidad de adecuar sus sistemas tecnológicos para hacer frente a la demanda de manera mas adecuada, enfocar bien a sus clientes y controlar los datos de su inventario. Esto permitiría a los vendedores tener una visión 360 grados de sus clientes y sus hábitos de compra a través de su multicanal.

Los sistemas operacionales estaban siendo usados para mejorar la inteligencia del negocio y obtener información de sus clientes. Esto había variado los grados de éxito, se incrementaban los problemas de calidad de información y tiempo de análisis.

Office Depot se dio cuenta que era hora de incrementar su inteligencia de negocio al siguiente nivel. Obtener una correcta herramienta y solución de *data warehouse* se volvió una prioridad. Esta herramienta debería ser capaz de ayudar a Office Depot a entender mejor su negocio, el cual es complejo por varias razones:



- Los multicanales por los que sus clientes son atendidos (catálogos, tiendas, internet y servicios de contrato)
- Los multiusos o destinos que le dan los clientes a los productos y servicios de Office Depot, ya que pueden ser comprados por un mismo cliente para su trabajo, familia, oficina de casa o actividades voluntarias.

Office Depot necesitaba una herramienta que le ayudara con las necesidades reales de los clientes al momento mismo de estar realizando una compra, por ejemplo, si un cliente estaba comprando una impresora, Office Depot quería ser capaz de ofrecerle productos relacionados como cables, cartuchos de impresión, sin importar el canal de compra que se estuviera utilizando.

La Solución

Office Depot seleccionó la solución analítica Teradata, adquirió Teradata Warehouse con 5 *terabytes*, y 4 *nodos*. Durante el primer año Office Depot incrementó el tamaño del *data warehouse* a 12 nodos debido a su actividad a través de la organización.

Los beneficios

Gracias a la implementación de este *data warehouse* Office Depot ha incrementado sus ventas y disminuido sus costos de inventario. El incremento en ventas viene a razón del profundo análisis de mercado y compra en sus múltiples canales, este análisis ha permitido un mejor entendimiento del comportamiento y hábitos de compra de los clientes, así como que productos se venden mejor en ventas sugeridas.

Tan solo en sus tiendas presenciales, la solución se pagó por sí sola en 1 año. El incremento en ventas gracias a las sugerencias de venta, se incrementó en dos dígitos desde la implantación de la solución. Actualmente la solución es usada por 1,200 usuarios con mas de 2,000 empleados recibiendo reportes de análisis de mercado, lo que les ha permitido un mejor entendimiento de los clientes.



4.2 Caso Delta Air Lines

Delta Air Lines es una aerolínea dedicada al transporte aéreo de pasajeros en Estados Unidos y gran parte del mundo, tan solo el año pasado, transportó 106 millones de pasajeros, lo que le produjo dividendos por \$15 billones, cuenta con 74,000 empleados a lo largo del mundo y 2,600 vuelos con destinos en todo lo ancho del planeta. Delta cree que su éxito se debe en gran medida a su superior servicio al cliente y que es definitivo, por lo que será su principal punto en el siguiente año. (NCR, 2001)

El reto

En una industria madura y de gran competencia, Delta trataba de encontrar la manera de acercarse a sus clientes diferenciarse a sí misma en el mercado (físico y web) y reducir sus costos. Pero con sus 27 *data marts* empleados y trabajando de manera independiente se ha dado cuenta que sus metas son todo un reto.

“En el pasado gastábamos mucha energía tratando de obtener números de diferentes sistemas analíticos para relacionarlos con los demás.” Explica Brent Lautenschlegar, director de servicios de *data warehouse* de Delta.

Con el sistema anterior, la estrategia existente de *data mart*, hacía la tarea de toma de decisiones mucho más difícil. Para aprovechar los datos operacionales y de los clientes que se tenían, Delta sabía que los usuarios necesitaban un fácil acceso a la información de toda la empresa.

La solución

Delta tenía los datos que necesitaba en sus sistemas, pero estaban fragmentados en diferentes áreas funcionales, los usuarios no podían ver tener una vista empresarial del negocio. La aerolínea rápidamente realizó una estrategia a lo largo de la empresa para satisfacer sus necesidades de información.

Delta centralizó los 27 *data marts* en un solo *data warehouse* para toda la organización con la ayuda de NCR, con su producto Teradata, lo que le permitió 2 terabytes de datos con capacidad para 5 terabytes. Rutinariamente la comunidad de usuarios de Delta empezó a hacer preguntas contra los datos detallados usando las herramientas SAS y tecnología analítica BRIO. Esto permitió a Delta optimizar rápidamente las oportunidades de negocio en toda la organización que permitió mejorar los niveles de servicio a sus clientes y mejorar la rentabilidad para sus accionistas.



Los beneficios

Gracias a la información más detallada que se tiene con Teradata de NCR, se ha podido incrementar la confianza en la calidad de información a la que pueden tener acceso y la velocidad de respuesta que anteriormente no era posible.

De acuerdo con el equipo de tecnología de información de Delta , es la velocidad de acceso lo que ha sido un gran factor para permitir al negocio ser mas productivo, Agregado claro a la mayor calidad de la información. Según Hyde "ahora tenemos mucha mas gente capaz de tomar decisiones importantes".

Con el ambiente computacional anterior, un requerimiento típico de información hacia el equipo de TI (tecnologías de Información), podría tardar semanas en tener un respuesta. Ahora los usuarios tienen mas respuestas en cuestión de minutos. Alta calidad en los datos y facilidad de lectura que permite a los usuarios tomar mejores y mas rápidas decisiones de negocio.

Otro gran beneficio fue la personalización hacia los clientes, en menos de un año, el *data warehouse* tuvo un tremendo impacto en el negocio de Delta. Tradicionalmente la industria de aerolíneas gratificaba a los clientes por la cantidad de millas que viajaban. Usando la información del *data warehouse* Delta cambió esa práctica relacionándola con la frecuencia de viaje, sin importar las millas. La aerolínea encontró que había baja relación entre los pasajeros que eran mas redituables y los que viajaban mas millas. Gracias a esa información, Delta pudo centrar su atención en aquellos valiosos voladores frecuentes y no solo a los que viajaban mas millas.

"Debemos tratar a nuestros pasajeros como individuos. El *data warehouse* provee herramientas para que podamos entender que tan individuales son cada uno de ellos y así podamos brindarles un mejor servicio." Dice Drewski. El *data warehouse* provee información para tratar a cada uno de los clientes por separado.

Delta también usa el *data warehouse* para mejorar la comprensión de ventas y canales de distribución, por ejemplo como compran los boletos (por agencia de viajes, por teléfono o por internet). Entendiendo esa mezcla y cómo cambios en áreas como notificaciones, practicas de boletaje e incentivos pueden afectar la manera en que los clientes compran los boletos. Delta puede mejorar su eficiencia y efectividad de su mercadotecnia y operaciones. Gracias al *data warehouse* Delta puede identificar fácilmente la mejor manera de venderle los boletos a sus clientes.



Capítulo 5.- Metodología de la investigación

Introducción

Este capítulo describirá la manera en que fue realizada la investigación de campo, para después mostrar como se diseñó la herramienta para la elaboración de dicha investigación y finalmente detallar los resultados obtenidos y su respectivo análisis.

5.1 Metodología

Esta tesis fue desarrollada basada en un método científico, lo cual nos permitirá asegurar que se ofrece un trabajo sustentado y con la validez suficiente para que pueda servir para trabajos futuros.

El trabajo realizado para la elaboración de este documento, ha tratado de ser lo más objetivo posible, dejando líneas de acción para futuras investigaciones y se ha tratado de que todo sea completamente verificable.

La investigación será de tipo descriptiva, ya que buscaré especificar y medir una propiedad importante de las empresas grandes en Monterrey, como es su nivel de implantación de sistemas de *data warehouse*. A través de los resultados de la investigación describiré el nivel promedio de avance, factores críticos de éxito, principales problemas y motivos principales en la implantación de estos sistemas en estas empresas.

La manera de realizar la investigación será de la siguiente forma: se formula el problema, se determina el objeto de la investigación, se reúnen los datos para analizarlos y posteriormente se emitirán conclusiones y recomendaciones.

Para la realización de esta investigación, se limitó la población a empresas grandes, ya que la implementación de un *data warehouse*, además de parecer una inversión fuerte, no es fácil de justificar para empresas medianas o pequeñas con una cantidad de información que no requiere de una herramienta de estas dimensiones.

La encuesta se aplicará principalmente a directores y/o gerentes del área de sistemas, ya que se considera que tienen una visión más clara de lo relacionado con los aspectos que nos interesa abordar con ésta investigación.



5.2 Herramienta de investigación (Encuesta)

Para la realización de esta investigación, se aplicó una encuesta (Ver Anexo 1) a un conjunto de empresas que representan el universo, para lo que se definieron previamente las preguntas de estudio.

Durante el desarrollo de la encuesta, se identificaron los aspectos importantes a considerar, de manera mas objetiva posible. De esta manera se determinaron las variables sobre las que se quería obtener información.

5.3 Variables de la investigación

Para la realización del instrumento de investigación, se identificaron algunas variables, las cuales servirían como base para el enfoque de las preguntas que se debían hacer a las personas encuestadas.

Estas variables, especifican lo que se quiere medir con la encuesta.

Variables dependientes:

- Nivel de avance en la implementación de un sistema de *data warehouse* en las empresas
- Factores críticos de éxito para la implementación de los sistemas de *data warehouse*
- Principales problemas en la implantación de un sistema de *data warehouse*
- Principales motivos para la implantación de un sistema de *data warehouse*

Las variables independientes son aquellas que nos ayudan a manipular el comportamiento de las variables dependientes. Es importante identificarlas, porque nos dan una idea del porque una variable dependiente se comporta de alguna manera en particular.

Variables independientes:

- Tipo de sistemas que se tenga instalado
- Tiempo que se lleve en la instalación
- Recursos de la empresa
- Etapa en que se encuentra la implantación



- Procedencia del sistema a implantar
- El tipo de empresa
- El involucramiento de la dirección
- Grado de necesidad de la información oportuna
- La cantidad de información de la empresa

Una vez identificadas las variables, se definieron las preguntas que conformarían la encuesta con la que obtendríamos la información.

5.4 Tamaño de la muestra

Durante la etapa de recolección de datos, se visitó el portal de SECOFI (www.secofi-siem.gob.mx/portalsiem/), donde se obtuvo el número de empresas grandes en la ciudad de Monterrey, las cuales representarán el universo de esta investigación.

Dentro de la página de SECOFI, se especifica que son consideradas *empresas grandes* aquellas empresas que cuentan con más de 500 empleados o que tienen unas ventas anuales mayores a 30 millones de pesos.

Una vez identificado el universo se identificó la muestra, la cual especificaría el número de empresas que se debían encuestar. La muestra fue el resultado de resolver la siguiente ecuación:

$$n = \frac{Z^2 * P(1-P) * N}{e^2(N-1) + Z^2 * P(1-P)}$$

Fuente: [Marcos, 1998]

Donde:

n= tamaño de la muestra

Z= nivel de confiabilidad; generalmente usamos 95%

P= probabilidad de éxito de que las empresas que se seleccionen tengan las características que se requieren

e= error esperado

N= tamaño de la población.



Sustituyendo las variables:

$$Z = .95$$

$$P = .95$$

$$e = .05$$

$$N = 80$$

Después de resolver la ecuación, se determinó que la muestra de esta investigación esta formada de 14 empresas.



Capítulo 6.- Resultados obtenidos y análisis

En el siguiente capítulo se mostrarán los resultados obtenidos de la aplicación del instrumento de investigación. Como se mencionó anteriormente, se realizaron 14 encuestas, a diversos tipos de empresas de la localidad entre las que se pueden mencionar Vitro, Banorte, John Deere, Cemex, Aerolitoral, Transportes Tamaulipas, Mabe, Banamex, Autobuses del Norte, Neoris y Regio limpieza, las otras 3 empresas pidieron confidencialidad, por lo que no son mencionadas. La distribución del tipo de empresas es de la siguiente manera:

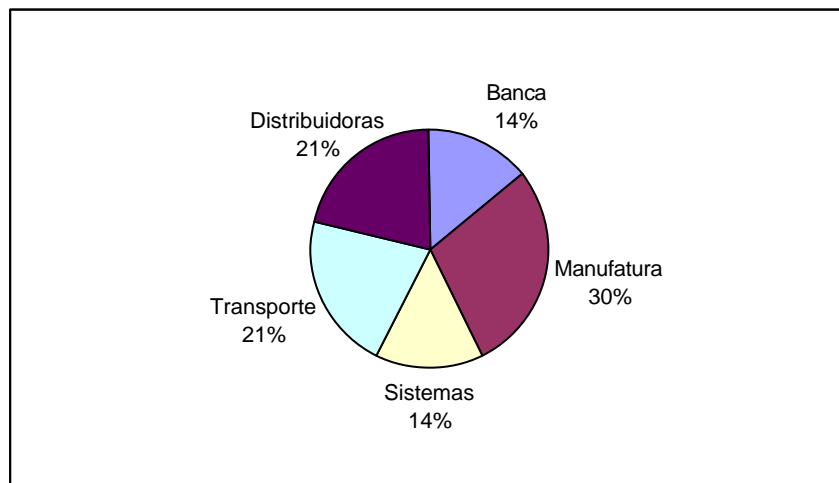


Fig. 6.1 : Distribución de las empresas por su giro

Los resultados de la investigación se presentarán de acuerdo con las variables de estudio que definieron la herramienta, y se presentará un breve análisis de las posibles causas de los resultados o de lo comentado por algunos de los entrevistados.



6.1 Información general

Para tener un marco de referencia más claro del tipo de empresas que conforman la muestra, a continuación se presentarán algunos aspectos de información general obtenida durante las entrevistas.

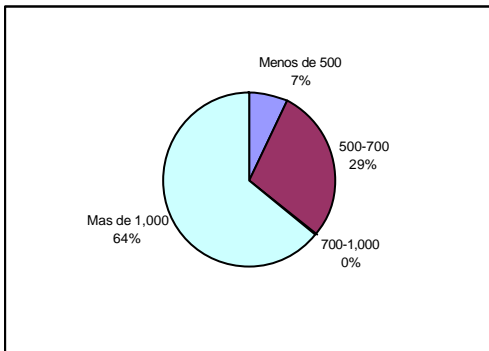


Fig. 6.2 Número de empleados

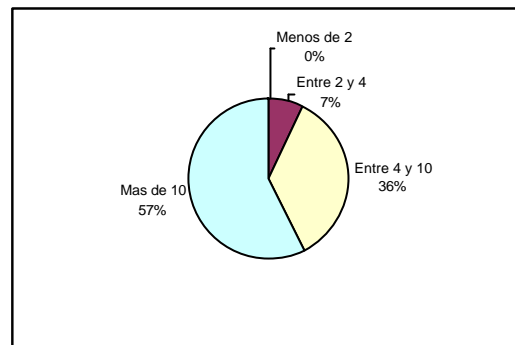


Fig. 6.3 Ventas anuales en millones de USD

Como podemos apreciar en las figuras 6.2 y 6.3, la mayoría de las empresas entrevistadas cuentan con más de 1,000 empleados y tienen ventas anuales por arriba de los 10 millones de dólares, mientras que el 29% de las entrevistadas tiene entre 500 y 700 empleados y el 35% vende entre 4 y 10 millones de dólares al año.

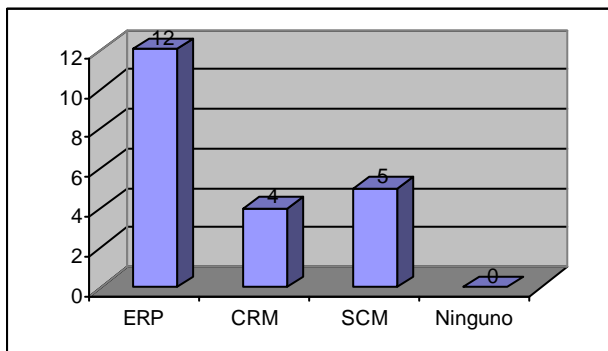


Fig. 6.4 TI con que cuentan

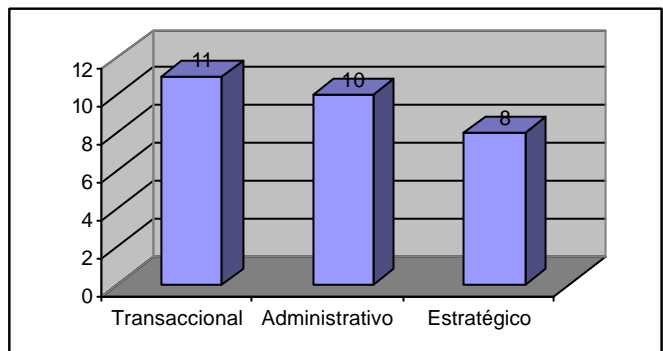


Fig. 6.5 Nivel que apoyan las TI

De las 14 empresas entrevistadas, 12 cuentan con tecnología de información que apoya al nivel transaccional mediante un ERP, una empresa que se dedica al desarrollo de sistemas cuenta con ERP, pero no considera que le brinde apoyo a nivel transaccional.

Por el tipo de empresas, 5 tienen administrado con tecnología de información su relación con los proveedores y solo 4 con los clientes. Considero que es bajo el



porcentaje de empresas que tienen tecnología de información apoyándolos al nivel estratégico, ya que es casi la mitad la que no cuenta con apoyo a este nivel.

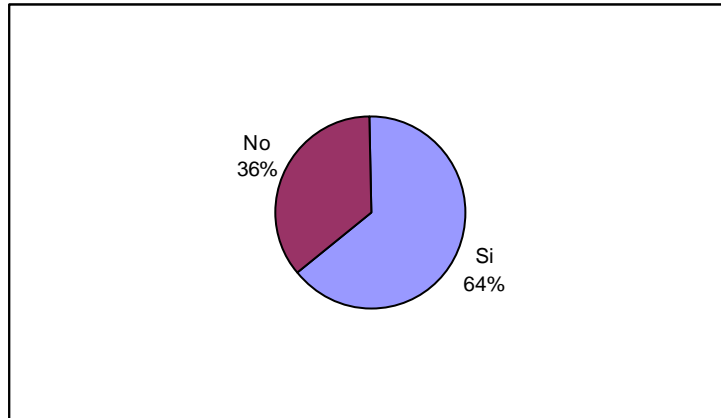


Fig. 6.6 Empresas que cuentan con *data warehouse*

El 64% de las empresas entrevistadas cuenta con alguna herramienta de *data warehouse*. La continuación de la encuesta se centrará únicamente en el 64% que sí cuenta con esta herramienta.

6.2 Nivel de implementación

El 100% de las empresas que cuentan con *data warehouse* consideraron que es importante la implementación de esta herramienta en la empresa, algunos de los entrevistados lo consideraron de vital importancia, como los relacionados con la banca, mientras que otros creen que apoya de manera eficiente sólo a los altos niveles de la organización.

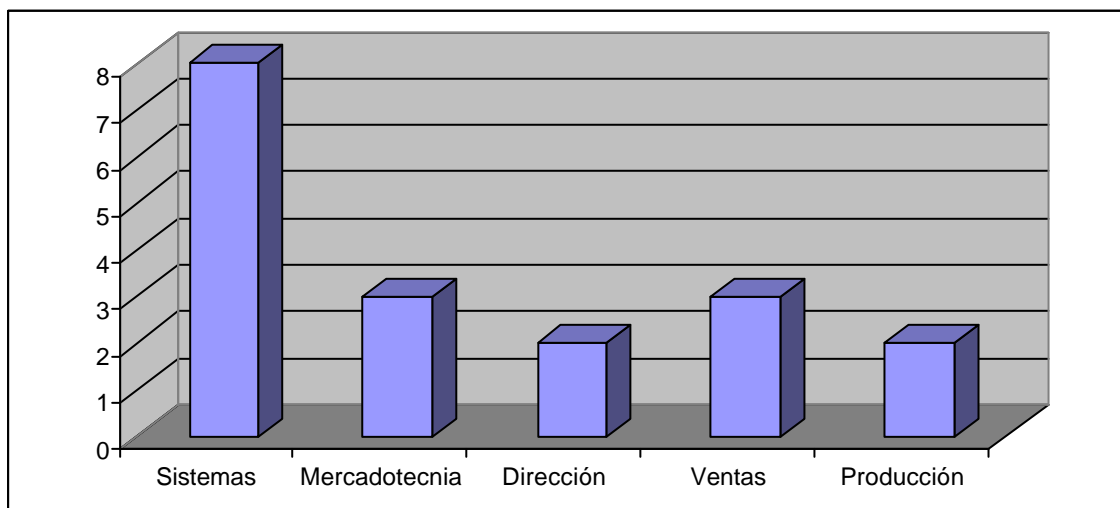


Fig. 6.7 Área que propuso la implementación del *data warehouse*



Como podemos apreciar en la figura 6.7, el departamento de sistemas sigue siendo el área que más se involucra en la implementación de un *data warehouse*, a pesar de ser una herramienta que no está dirigida al soporte de esta área. Así mismo, se puede apreciar que la dirección es el área que menos propuso la implementación, esto nos da una idea de la cultura computacional que rodea a las empresas de la entidad, se explica gráficamente el que a veces se piden cosas sin importar o tener conocimiento de los medios necesarios para llegar a lo que se esta pidiendo.

Además de las áreas listadas en la gráfica, estaban como opciones las áreas de recursos humanos y finanzas, sin embargo no obtuvieron ninguna ocurrencia, por lo que no fueron incluidas en la gráfica.

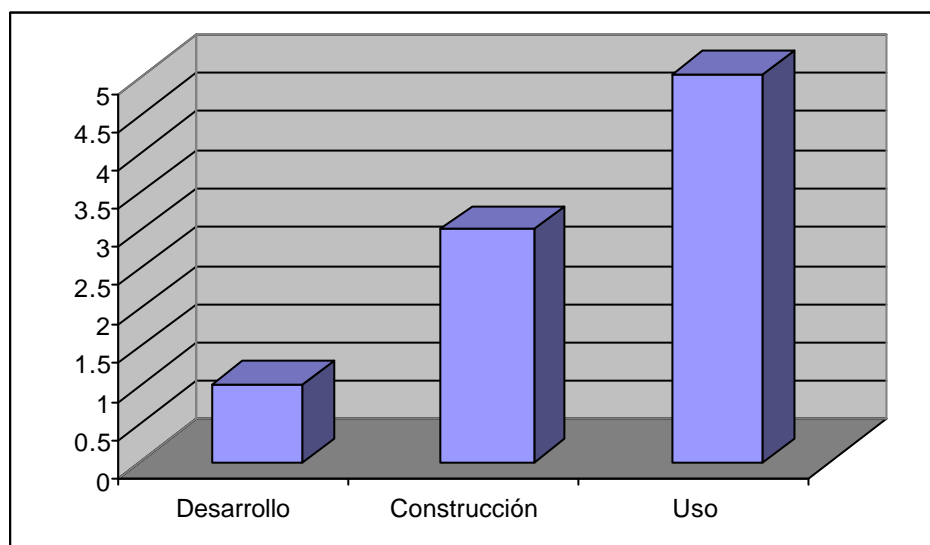


Fig. 6.8 Fase de implantación en que se encuentra la herramienta

Como se puede apreciar en la figura 6.8, el 56% de las empresas entrevistadas ya están en la fase de uso de la herramienta, dentro de estas empresas se encuentran principalmente las relacionadas con la banca y transporte aéreo, en las cuales ya se tiene un considerable tiempo en la utilización de la herramienta y la consideran imprescindible para el desarrollo de sus operaciones diarias. Las empresas en las que la herramienta se encuentra en desarrollo y construcción, comentaron haberse dado cuenta de la necesidad de contar con este tipo de herramienta, y están trabajando en habilitarla lo antes posible.

Cabe mencionar que en la encuesta no se ofrecía como opción la fase de "uso", pero todas las que incurrieron en esta fase, eligieron "otra" y especificaron que se encontraba en uso.

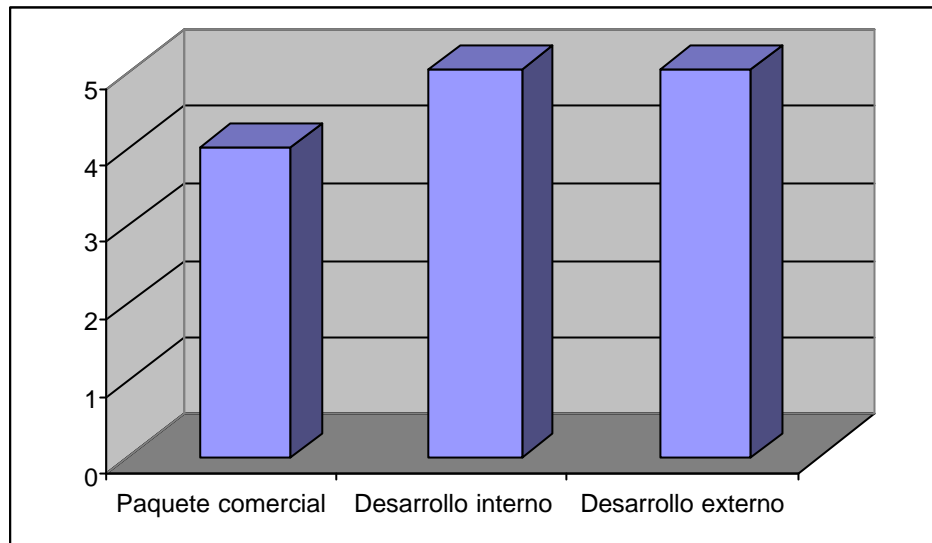


Fig. 6.9 Origen del software de data warehouse

Solamente una de las cuatro empresas que compraron un paquete comercial, no requirieron de adecuaciones, ya sea por parte de desarrollo interno o mediante consultores que lo desarrollaran externamente. El restante 88% de las empresas, compraron un paquete y tuvieron que adecuarlo a su giro o procesos internos para poder obtener un mejor beneficio de la herramienta. Las empresas están concientes de que ningún paquete comercial cumplirá exactamente con sus necesidades de negocio, y están dispuestas a pagar su dependencia con paquetes comerciales a pesar de que tienen que buscar otras soluciones por su lado.

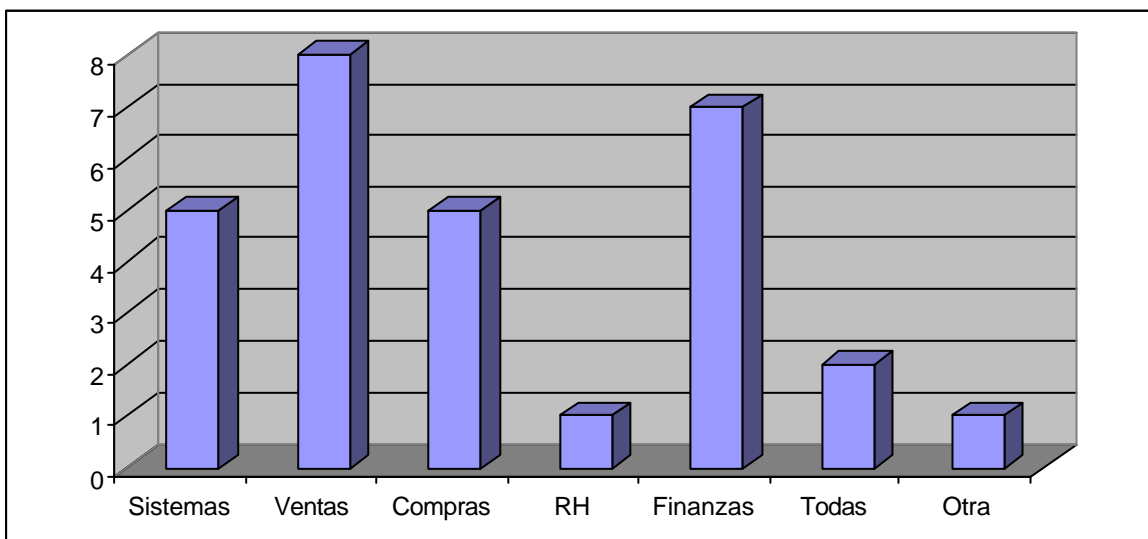


Fig. 6.10 Áreas contempladas por el data warehouse



Las áreas de ventas, finanzas y compras son las más involucradas con el uso de la herramienta. A pesar de que la información de todas las áreas debe contemplarse en el *data warehouse*, son las áreas de ventas, finanzas y compras las que generan información más relevante y de más peso para el apoyo a la toma de decisiones, por lo que tienen una prioridad más alta para ser contempladas por la herramienta.

La empresa que contestó que otra área estaba involucrada, especificó que se trataba del área de producción, ya que por su tipo de negocio era de vital importancia su involucramiento para la generación de información.

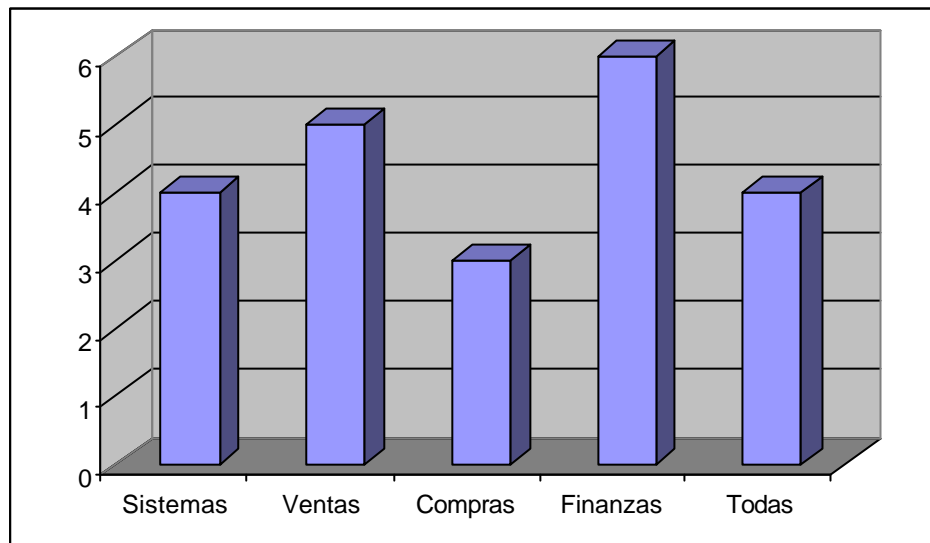


Fig. 6.11 Áreas más beneficiadas con el *data warehouse*

Todos los entrevistados concluyeron que todas las áreas son beneficiadas, sin embargo se hizo especial énfasis en que especificaran cuales áreas obtenían mejor beneficio o reflejaban con hechos mas palpables lo que obtenían de la herramienta, y fue entonces cuando el 67% dijo que el área de finanzas era la mas beneficiada, mientras que el 56% coincidió que es ventas de las que más se apoyan.

La implementación de todo sistema o herramienta computacional debe reflejar o apoyar a las áreas que tengan más que ver con la generación y la administración del capital, ya que esto hace más fácil la justificación de una inversión de estas magnitudes.

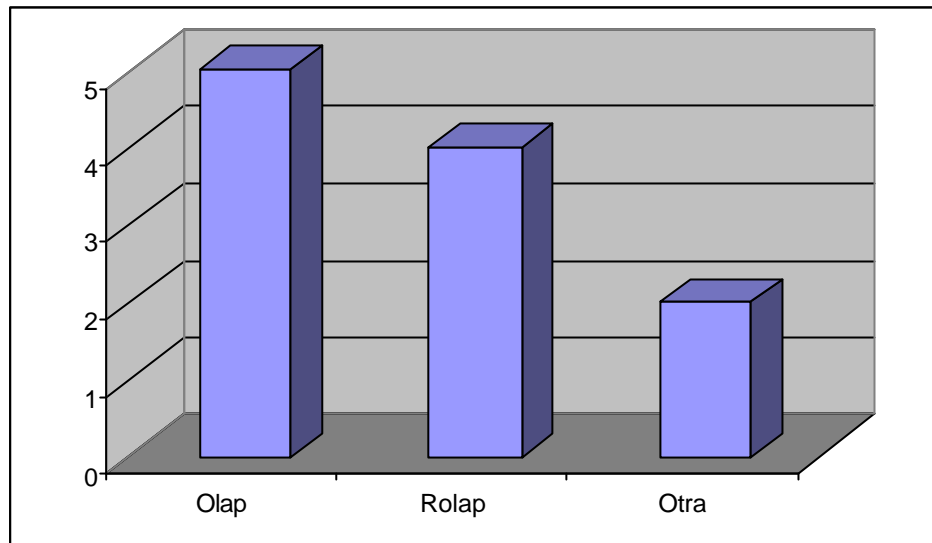


Fig. 6.12 Herramientas de explotación usadas

Como se puede apreciar en la figura 5.12, las técnicas más usadas para la explotación del *data warehouse* son *Olap* y *Rolap*, las 2 empresas que dijeron usar otro método, usaban también *Olap*, y explicaron que usaban métodos internos para extraer información del recipiente y para usarlos en procesos muy específicos.

Como vimos en capítulos anteriores, según Fernández (1999), la técnica *Olap*, tiene entre sus principales características soportar requerimientos complejos de análisis, analizar datos desde diferentes perspectivas, así como, soportar análisis complejos contra un volumen grande de datos. Este tipo de técnica es el usado por empresas con un volumen de información inmanejable para sistemas independientes. Mientras que los que usan *Rolap*, se inclinan por esta técnica principalmente por el tipo de base relacionales con las que cuentan.

Además de los presentados, la encuesta ofrecía opciones como *Datamining*, *Molap* y *Webhousing*, sin embargo no tuvieron incidencia alguna, por lo que no son presentados en la figura 6.12.

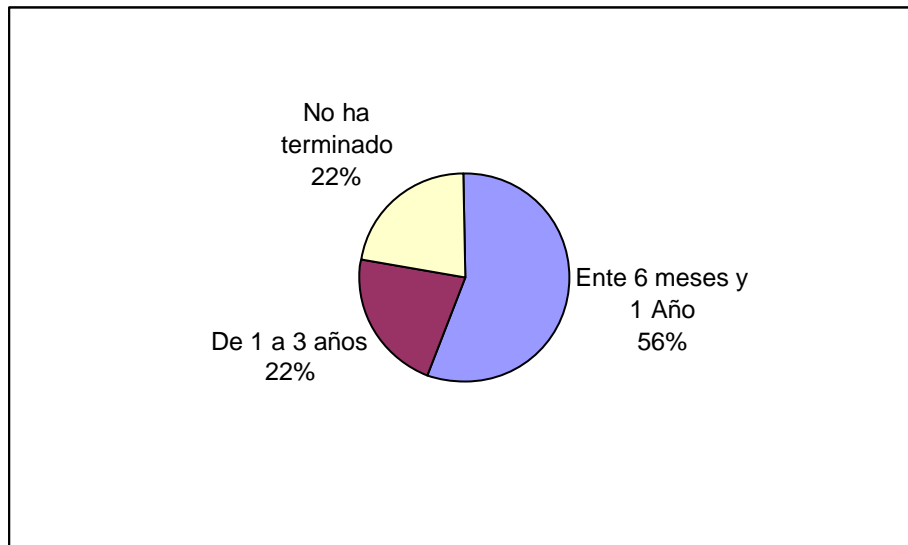


Fig. 6.13 Tiempo que duró la implementación

Como se muestra en la figura 6.13, el 56% de las empresas implementaron la herramienta en menos de un año, a pesar de que requerían hacerle adecuaciones al paquete comercial comprado en la mayoría de ellas, estaban concientes de la necesidad de contar con la herramienta funcionando en todo su potencial por lo que dedicaron grandes esfuerzos y cantidades de dinero para asegurarse de terminar lo antes posible.

Entre los entrevistados que contestaron que no habían terminado, uno de ellos comentó que llevan 3 años en el proyecto, mientras que el otro llevaban apenas 6 meses con el inicio de la implementación.

6.3 Motivos para la implementación

Las siguientes preguntas estuvieron encaminadas a determinar los principales motivos para la implementación de un *data warehouse*

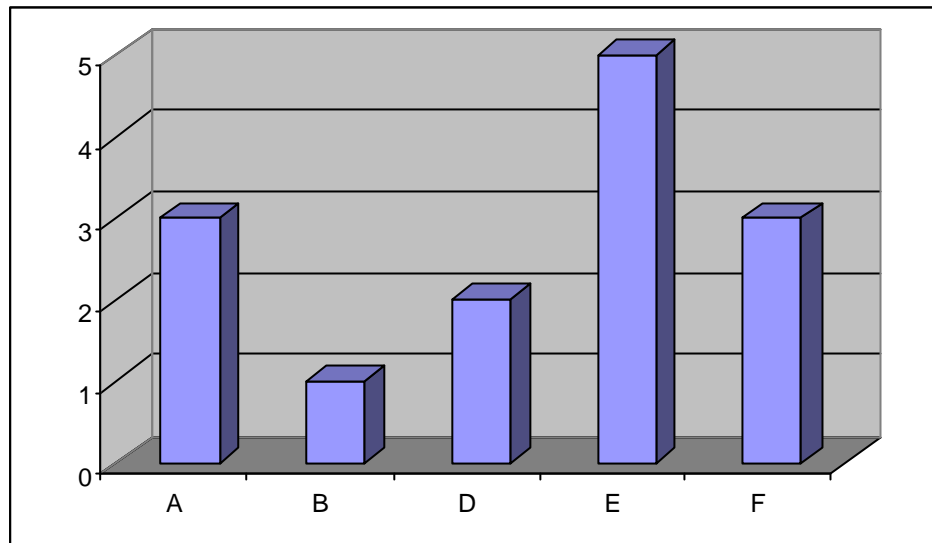


Fig. 6.14 Elementos o variables que son necesarios identificar en una empresa para considerar la implementación de un *data warehouse*

En la encuesta, las letras de la figura 6.14 tenían los siguientes valores:

- A) Administración actual de los sistemas transaccionales.
- B) Presupuesto para invertir en el proyecto
- C) Personal especializado en la gestión de *data warehouse*
- D) Estrategias de negocio orientadas a inversión en tecnologías de almacenamiento masivo de datos
- E) Todas
- F) Otra

Como podemos apreciar, la mayoría de las empresas entrevistadas coincidieron en que todos los aspectos eran importantes, sin embargo consideraban que uno de los puntos más importantes era la necesidad de administrar la información generada por los sistemas transaccionales.

Entre las respuestas no consideradas por la encuesta pero si por los entrevistados tenemos la administración del negocio, agilización para la toma de decisiones e identificar los indicadores del negocio que se quieren medir.

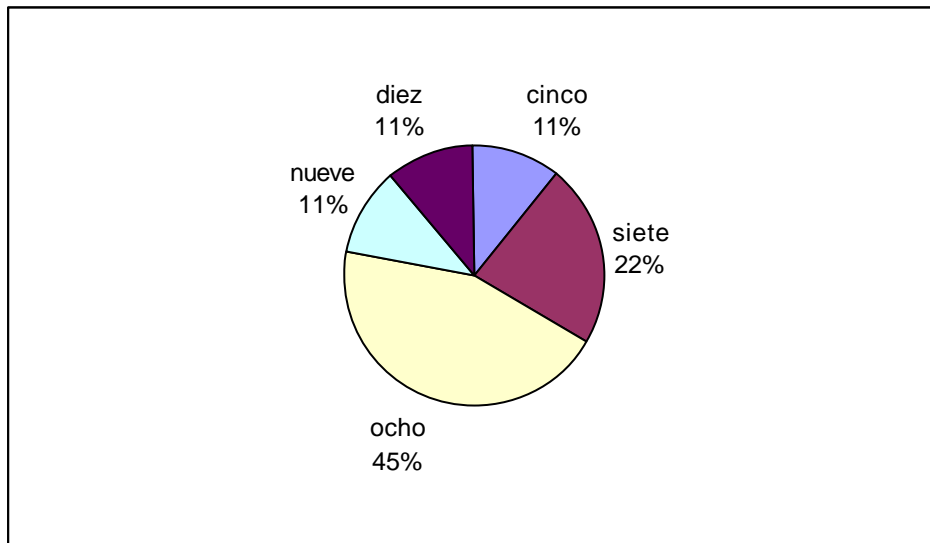


Fig. 6.15 Calificación brindada a la productividad operativa generada por el *data warehouse*

En una escala de 1 a 10, siendo 1 nada y 10 Mucha, la figura 6.15 nos muestra la manera en que los entrevistados calificaron la productividad operativa que ofrecía la herramienta de *data warehouse*. La persona que calificó con cinco es del giro bancario, y explicó que el *data warehouse* no estaba tan enfocado al cajero, por lo que no representaba realmente una ventaja operativa para ellos, estaba más enfocado a niveles más altos dentro de la organización.

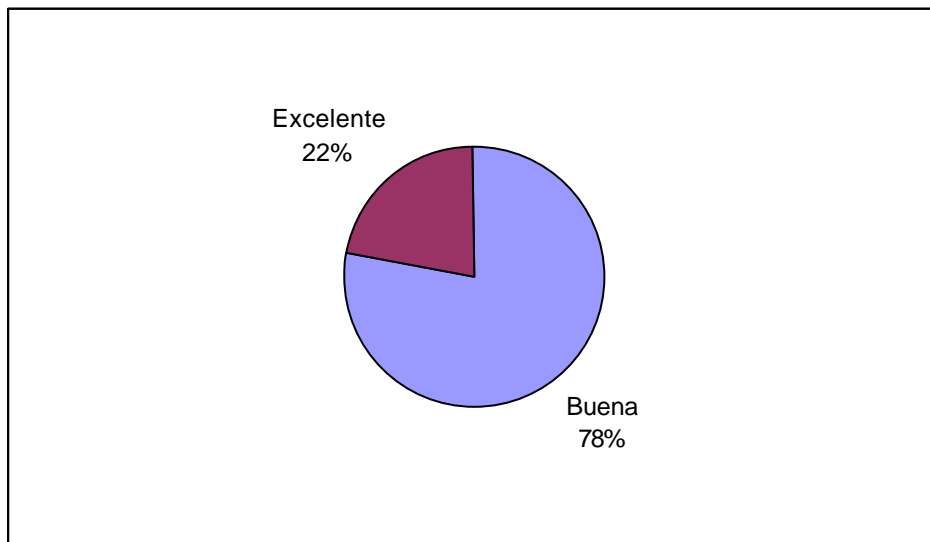


Fig. 6.16 Apreciación de la relación costo-beneficio al implementar *data warehouse*

La figura 6.16 muestra lo que las empresas que cuentan con una herramienta de *data warehouse* perciben al tratar de hacer un estudio costo-beneficio de la implantación



de esta herramienta. Un 78% identificaron como *buena* la relación, mientras que el restante 22% como *excelente*, lo cual habla bien de los dividendos que se obtienen gracias a la herramienta en caso de saber aprovecharla de manera adecuada.

El 100% de los encuestados coincidieron en que la implementación de un *data warehouse* en la empresa ha mejorado de alguna manera la eficiencia en la toma de decisiones de las personas encargadas de esta función. Es importante mencionar que el contar con la herramienta en sí, no cambia nada para el negocio, mucho tiene que ver la forma en que las personas que usan esa información, la asimilan, entienden y usan en pro de elegir mejores opciones alineadas a las estrategias de la empresa.

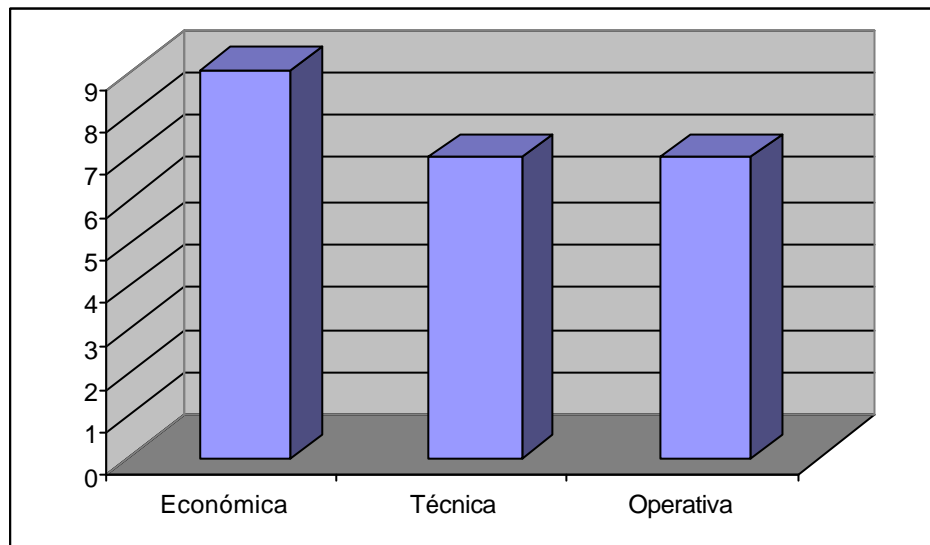


Fig. 6.17, Estudios de factibilidad realizados para la adquisición del *data warehouse*

El 100% de los entrevistados analizaron diferentes proveedores antes de adquirir el paquete con el que cuentan ahora, entre los que comentaron que descartaron están IBM, SAGA y Compaq.

Así mismo, todos hicieron estudios de factibilidad económica, en cambio, solo el 78% hicieron estudios de factibilidad técnica y operativa, lo que hace pensar que aquellos que no los hicieron, resolvieron adquirir una herramienta de este tipo basados principalmente en base al precio.

La factibilidad técnica y operativa de un proyecto debe ser uno de los puntos más importantes, ya que evita muchos dolores de cabeza principalmente a la hora de la implementación y de la operación de la misma, el no realizar estos estudios puede hacer que nos demos cuenta demasiado tarde de que adquirimos algo que no va a resolver ninguno de nuestros problemas o que tendremos que cambiar mucho de nuestra infraestructura operativa o tecnológica para que proporcione los resultados esperados.

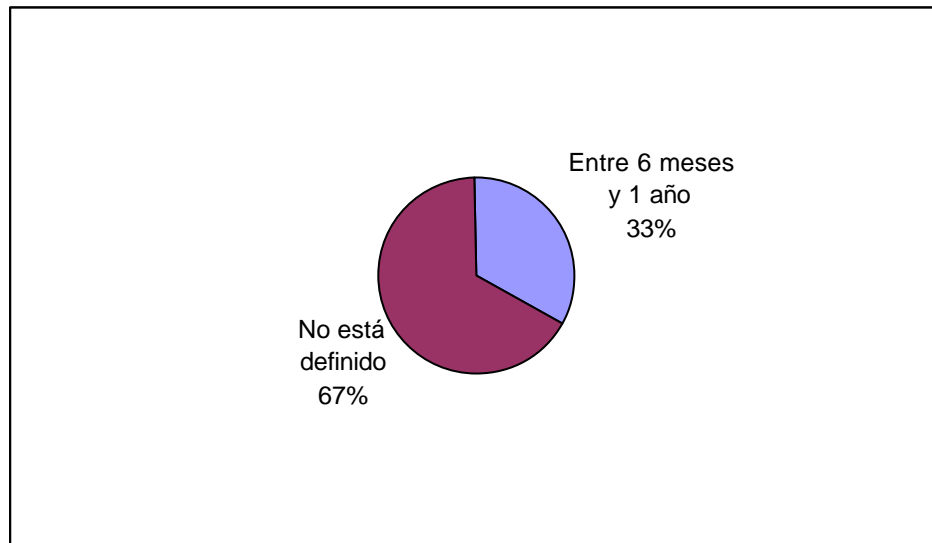


Fig. 6.18 Tiempo estimado para el retorno de inversión de la implementación de un *data warehouse*

Todas las empresas coincidieron que era difícil estimar el tiempo en el que recuperarían la inversión, ya que los beneficios son cualitativos y no cuantitativos, lo que hace que dependa mucho del tipo de decisiones que permita tomar, a lo que es difícil aplicar un valor monetario. Sin embargo el 33% estimó que entre 6 meses y 1 año recuperarían lo invertido en la herramienta. Los estudios de factibilidad económica fueron centrados principalmente en la capacidad de adquirir la herramienta y la comparación con los demás proveedores.

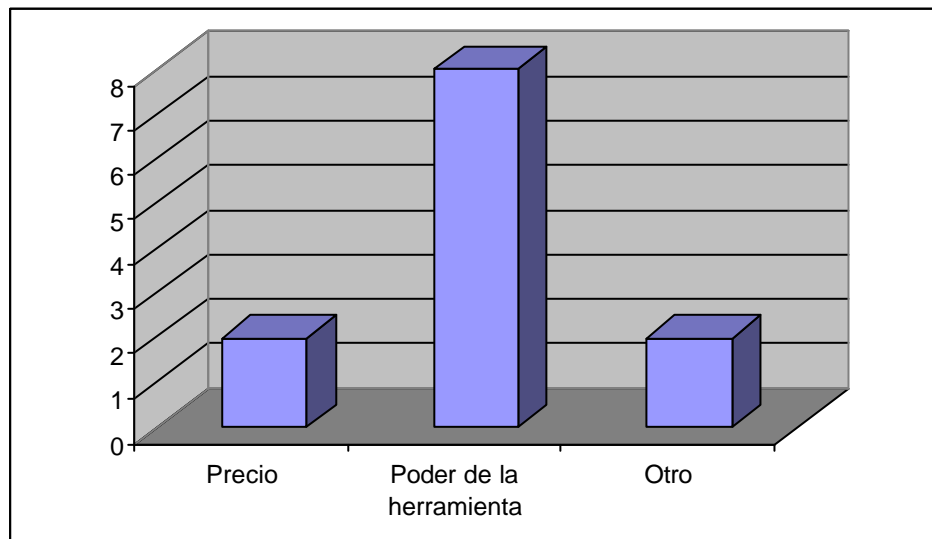


Fig. 6.19 Factores de decisión para adquirir la herramienta actual



Como podemos apreciar en la figura 6.19, 89% de las empresas tomaron en cuenta el *poder de la herramienta* como factor clave para decidir cual comprar, 22% incluyeron el *precio* dentro de los factores determinantes, y otro 22% incluyeron otro factor entre los que encontré *consultores que avalan la exitosa implementación de la herramienta y el tiempo de implementación*.

Entre los factores que se propusieron en la encuesta y que nadie eligió encontramos:

- Experiencias anteriores
- Recomendación de 3eros
- Requerimientos de hardware

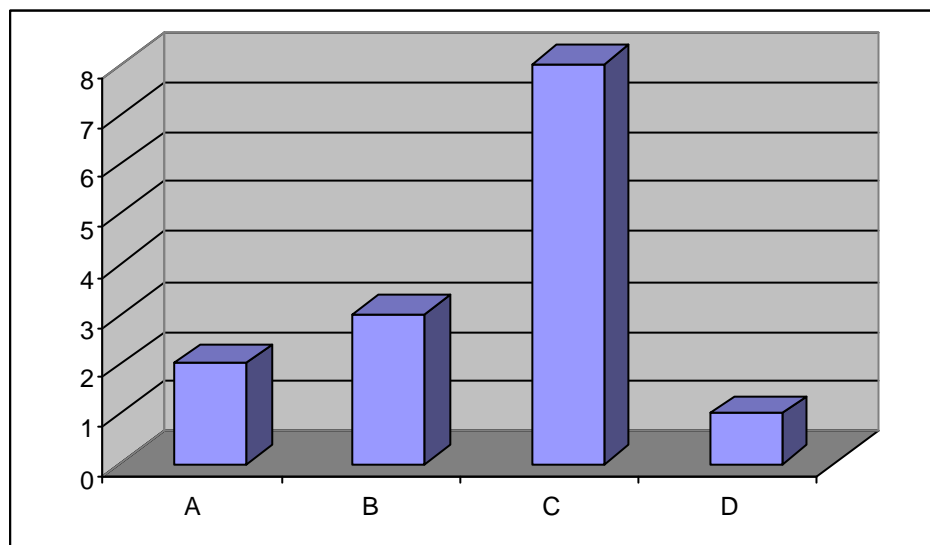


Fig. 6.20 Principales motivos para la implementación de una herramienta de *data warehouse*

En la figura 6.20 los valores de las letras en la encuesta son:

- A) Centralización de información en un repositorio único de los sistemas transaccionales
- B) Depuración de la información transaccional
- C) Necesidad de análisis y toma de decisiones sobre grandes volúmenes de información
- D) Se presentó la oportunidad en términos de costo de herramientas y consultoría.

Como era de esperarse, el 89% consideró la necesidad de análisis y toma de decisiones sobre grandes volúmenes de información como uno de los principales motivos para la implementación, esto parece lógico, ya que es una de las características de más peso del *data warehouse*. Un 33% consideró la depuración de



su información transaccional como un factor importante, lo cual va muy relacionado con el factor antes mencionado.

Parece curioso, pero solo 1 de las 9 empresas que cuentan con esta herramienta incluyó finalmente el costo como un factor determinante para la implementación.

6.4 Factores críticos de éxito en la implementación

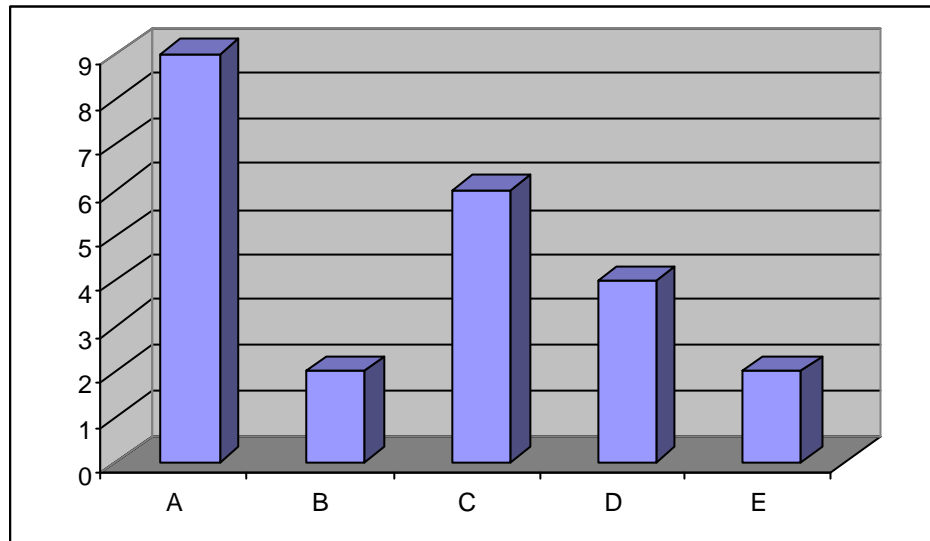


Fig. 6.21 Variables o elementos claves para realiza una implementación exitosa

Los valores en la encuesta de las letras de la figura 5.21 son:

- A) Planeación y administración del proyecto manejada por fases, tomando las variables del negocio como unidades independientes (*Data Mart*)
- B) Planeación y administración del proyecto manejada por fases, tomando las variables del negocio como una sola unidad (*Data Warehouse*)
- C) Definición clara de donde se encuentra la información que almacenará el *data warehouse*, las reglas de negocio y los indicadores de desempeño.
- D) Diseño del repositorio de datos que permite explotar la información de manera multidimensional para alcanzar las capacidades de análisis esperadas.
- E) Personal con conocimientos de uso y manejo de arquitectura de *data warehouse*

Como se puede apreciar en la figura 6.21, el 100% de las empresas consideró un factor de éxito la planeación y administración del proyecto manejada por fases, tomando cada unidad de negocio como algo independiente. El 67% coincidió que es importante identificar la ubicación de la información, las reglas del negocio y los indicadores de desempeño. Comentando con los entrevistados, también se definió como factor de éxito, el tener los objetivos del proyecto bien claros y alineados con las estrategias del negocio.

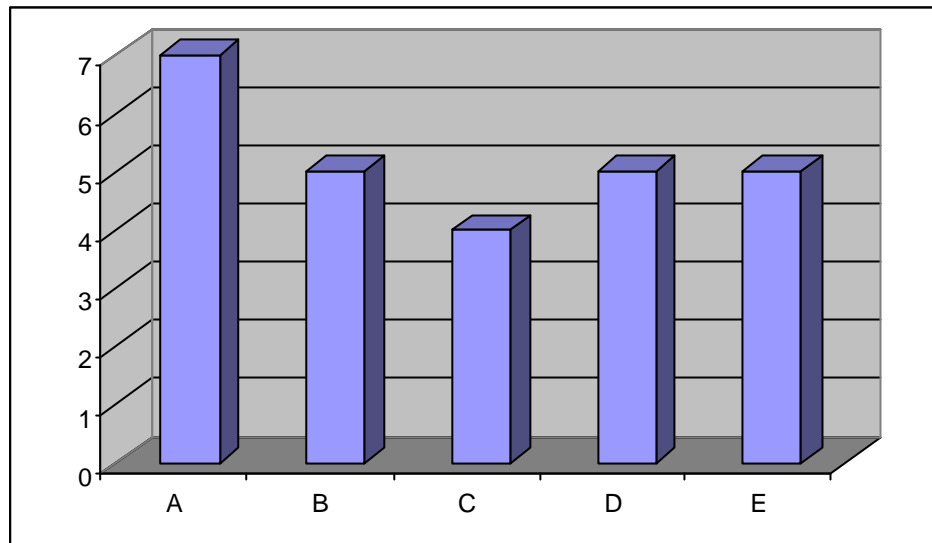


Fig. 6.22 Motivos de éxito de *data warehouse*

Los valores en la encuesta de las letras de la figura 6.22 son:

- A) Alineación entre los requerimientos de información ejecutiva y el trabajo del personal de sistemas.
- B) Efectiva extracción, transformación y carga de los sistemas transaccionales.
- C) Herramientas tecnológicas adecuadas que soporten la infraestructura planteada y los volúmenes de información a manejar.
- D) Coordinación entre el diferente personal involucrado en el proyecto.
- E) Cultura de uso del *data warehouse* por parte de los gerentes y directores.

Entre los principales motivos que las personas encuestadas eligieron para considerar su *data warehouse* un caso de éxito, encontramos que el 78% lo atribuye a la alineación ente los requerimientos de información ejecutiva y el trabajo del personal de sistemas. Se hizo énfasis en el involucramiento de los directivos y las personas que usan la información para la toma de decisiones, sin embargo solo para el 56% de los encuestados representó éste un motivo de éxito.

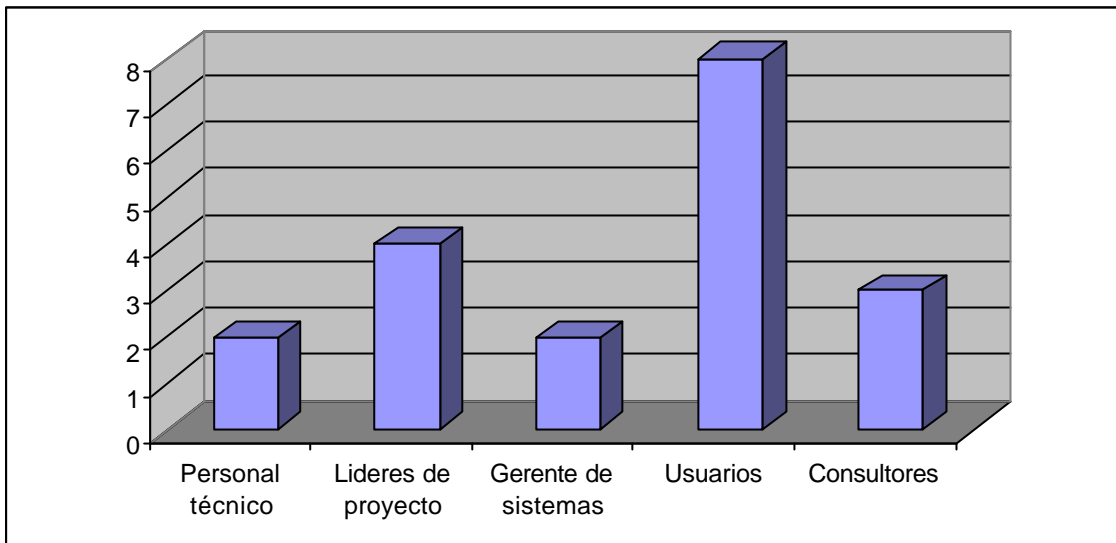


Fig. 6.23 Involucrados en el éxito de la herramienta

Es un hecho que para tener un *data warehouse* exitoso, se requiere el involucramiento de mucha gente, el 89% de las personas encuestadas coincidieron en que el usuario es el principal propiciador de este éxito, seguido por los líderes de proyecto, ya que son quienes se deben encargar de los tiempos y objetivos del mismo, después un adecuado respaldo de consultores externos, que guíen y soporten al líder, y por último los gerentes o encargados del área de sistemas y el personal técnico.

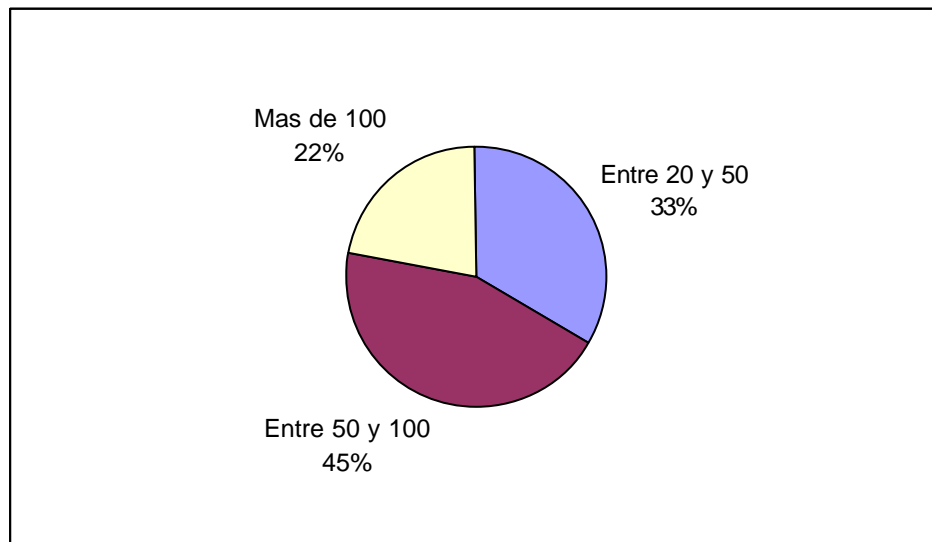


Fig. 6.24 Monto aproximado de la inversión en *data warehouse* en miles de dólares



Como se puede apreciar en la figura 6.24, el 45% de los encuestados invirtieron entre 50 y 100 mil dólares en la inversión de su *data warehouse*, 33% invirtieron entre 20 y 50 mientras que el 22% gastó más de 100 mil dólares. Según comentaron los encuestados, el monto variará directamente con la marca del *software*, además de la rapidez con que se desee implementar y las áreas que se quiera involucrar. Para empresas del tipo de la muestra, una inversión de estas magnitudes no representa un gran peso, siempre y cuando estén bien respaldadas con un estudio de la utilidad que la herramienta traerá, así como la importancia que tiene la misma en las operaciones diarias, por ejemplo en la banca, este tipo de tecnología es de vital importancia, ya que sin esta no se pueden realizar las operaciones más triviales del nivel transaccional.

Entre los aspectos considerados por el monto elegido, los entrevistados incluyeron:

- Costo del paquete
- Capacitación a los usuarios
- Ayuda de consultores, en los casos necesarios
- Horas del personal interno del área de sistemas para la adecuación de la herramienta a las necesidades del negocio
- Hardware en caso de no contar con él
- El tiempo de los usuarios para evaluar la herramienta

6.5 Problemas más comunes

Por último, esta sección de preguntas trató de identificar que problemas encuentran las empresas al implementar una herramienta de *data warehouse*.

El 100% de los encuestados admitieron haber tenido algún tipo de problema al implementar su *software*, todos coincidieron también en que los problemas que enfrentaron no eran privativos de su empresa ni de su giro de empresas, es decir, que eran problemas que ellos consideraban que se le podían presentar a cualquiera.

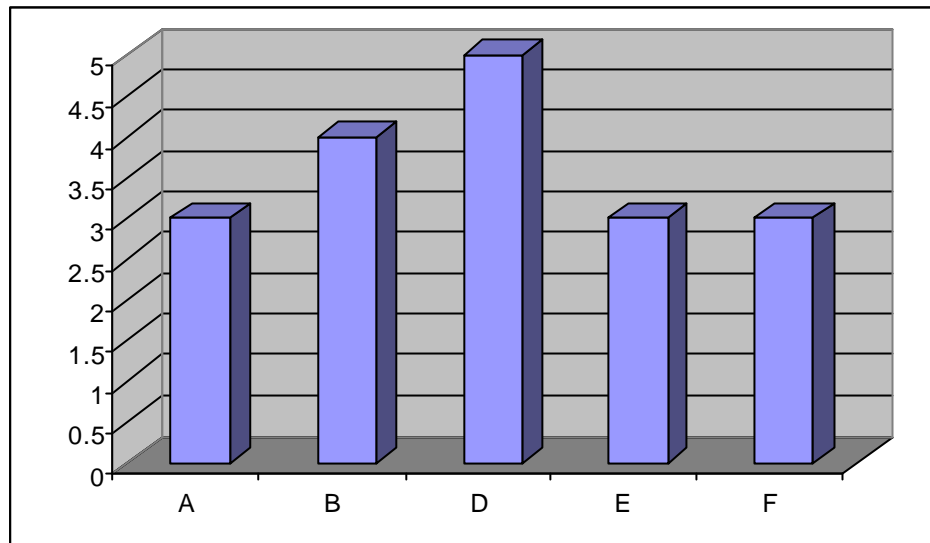


Fig. 6.25, Problemas que enfrentaron el implementar la herramienta

Los valores en la encuesta de las letras de la figura 6.25 son:

- A) Resistencia a la implementación por parte del personal o usuarios
- B) Falta de involucramiento por parte de los directivos
- C) El hardware no cumplió con las necesidades del software
- D) No se cumplían las fechas estipuladas
- E) Problemas relacionados con la falta de capacitación o experiencia por parte de los implantadores
- F) Otro _____

En la implementación del 56% de los encuestados tuvieron problemas con el cumplimiento de las fechas por parte del personal de sistemas y de los consultores para terminar las fases de implementación. En el 44% de las empresas los directivos no parecían involucrados, lo cual obstaculizaba el accionar del personal que tenía que realizar la implementación, sin embargo varios encuestados comentaron que al ver la falta de cumplimiento de fechas, los directivos ponían más atención en el proyecto lo que permitió su mejor involucramiento y dio agilidad al mismo.

Entre los problemas que se encontraron las empresas no considerados por la encuesta podemos encontrar:

- Problemas relacionados con la manera en que se encuentran almacenados actualmente los datos y problemas para ajustar la herramienta a la infraestructura tecnológica del negocio.
- Problemas de integridad de la información.
- Los datos se encuentran en formatos que hacen difícil la extracción y carga, además este tipo de sistemas rompen paradigmas y mucha gente,



principalmente usuarios no pueden ver las ventajas del uso de la información multidimensional.

La última pregunta del cuestionario hacía referencia a los problemas que los encuestados consideraban más importantes o que consideraban que más obstaculizaban la implementación de un *data warehouse*, la respuesta podemos verla en la figura 6.26.

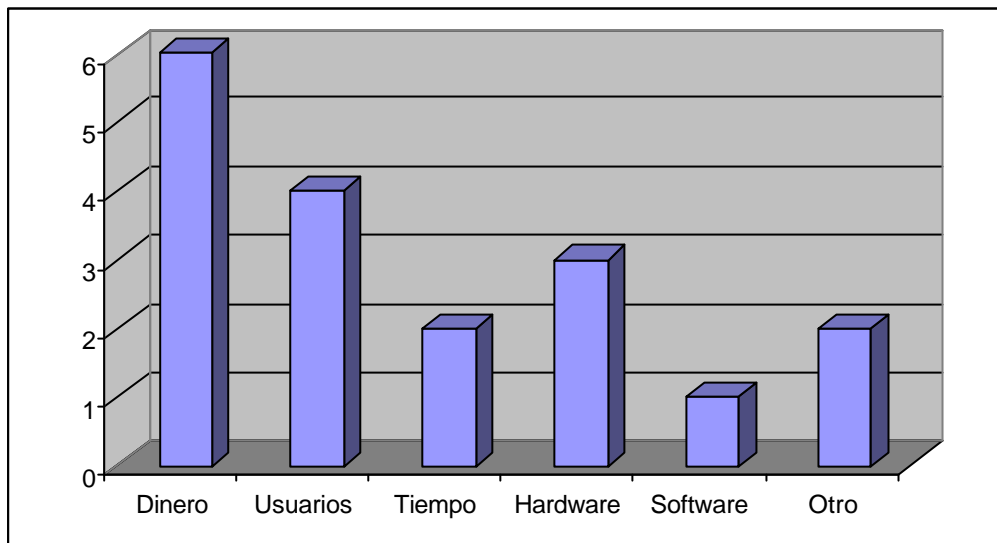


Fig. 6.26 Problemas más importantes o que más obstaculizan la implementación de un *data warehouse*

Como podemos ver en la figura 6.26, el 67% de los encuestados puso a los aspectos relacionados con el *dinero* como el problema más importante para la implementación de la herramienta, *los usuarios* fueron considerados de los mayores problemas por el 44% de la muestra, el software fue el que tuvo menos ocurrencias con tan solo 11% de las opiniones, cabe mencionar que nadie consideró a la dirección general como un problema a pesar de que era considerada en la encuesta como una de las opciones.

Dentro de los que consideraron otro problema además de los listados en la encuesta, podemos mencionar:

- La balanza entre tiempo, requerimientos y administración del proyecto.
- Encontrar al proveedor del software que mejor se adapte a las necesidades de la empresa
- Cumplimiento de los proveedores



Se consideró a los aspectos relacionados con el dinero como uno de los problemas, debido a lo difícil que es calcular el monto total de la inversión, ya que la compra de un paquete comercial es solo el inicio de una gran inversión que se ve incrementada por horas de capacitación, requerimientos de hardware, tiempo del personal de sistemas, ya sea interno o consultores, que deberá afinar los detalles de la herramienta de manera que se le pueda sacar el mayor provecho, etc., lo cual hace difícil el cálculo del retorno de inversión, por lo que es importante sustentar bien el proyecto en base a los beneficios que se esperan obtener y casos de éxito en empresas del mismo giro.



Capítulo 7.- Conclusiones, recomendaciones y trabajos futuros

7.1 Conclusiones

En base a los resultados de las encuestas aplicadas a las diferentes empresas, lo citado en el marco teórico y a mi experiencia a través de la maestría, se puede concluir que el uso de las herramientas de *data warehouse* son un área de oportunidad para muchas de las empresas grandes de la ciudad de Monterrey, ya que solo un poco más de la mitad de las empresas cuentan con esta herramienta, y sin embargo está completamente demostrada su utilidad y capacidad.

La aplicación efectiva y eficiente de una herramienta de esta naturaleza puede marcar la diferencia entre una empresa exitosa y una del montón, ya que el contar con la información correcta, completa y oportuna es una ventaja competitiva que cada vez toma más importancia en un mundo globalizado donde la información se vuelve cada vez más un factor de diferenciación.

Data warehouse parece una buena opción para aquellas empresas con un gran volumen de datos que necesitan contar con administración y acceso rápido y digerido de la información que sus sistemas transaccionales generan.

Por otro lado, se puede identificar que todas las empresas que contestaron que contaban con tecnología de información apoyando al nivel estratégico, cuentan con una herramienta de *data warehouse*, por lo que es fácil concluir que el contar con la información organizada de manera que un agente de búsqueda pueda establecer relaciones o patrones entre la información de toda la organización, permite que el nivel estratégico cuente con información que no es fácil visualizar a simple vista.

El área de sistemas sigue subiendo en el escalafón empresarial para la toma de decisiones importantes que afectan radicalmente la manera en que las organizaciones realizan sus procesos críticos o están organizadas, por lo que existe actualmente mucha oportunidad para egresados de esta área y en especial para aquellos que cuentan con una especialidad en administración de tecnologías de información ya que los retos de competencia actuales hacen que las empresas busquen lo mejor del capital humano para estar a la vanguardia y aprovechar los recursos que la tecnología ofrece para hacer más rentable el negocio y así incrementar la productividad de toda la organización, invirtiendo más en tecnología que mejore y agilice los procesos en vez de en personal que duplique las horas de trabajo o realice más actividades en la jornada laboral.



7.1.1 Nivel de implementación

En base a los resultados obtenidos en las encuestas, se puede considerar que el nivel de avance en la implementación de *data warehouse* en las empresas en Monterrey está por debajo del nivel en que debería estar, principalmente por tratarse de una de las ciudades con mayor población de empresas grandes en el país, lo que debería implicar que se contara con un avance en tecnología aceptable, sin embargo vemos como apenas la mitad de la población cuenta con un *data warehouse* en uso, mientras que las demás empresas lo están construyendo, conceptualizando, o ni siquiera lo tienen, lo cual debe preocupar a aquellos directivos de estas empresas en caso de querer continuar en el mercado compitiendo con empresas internacionales de su ramo que si pudieran contar con la información que ellos no están aprovechando.

7.1.2 Factores críticos de éxito

Por otro lado, con respecto a los principales factores críticos de éxito para la implementación de los sistemas de *data warehouse*, se puede concluir que se debe:

- Dividir el proyecto por unidades independientes del negocio
- Identificar la información que se desea almacenar de manera correcta y "limpia"
- Identificar las reglas del negocio e indicadores de desempeño
- Un buen diseño multidimensional del repositorio de datos para su efectiva explotación
- Involucramiento de la dirección y usuarios en general
- Objetivos de la implantación bien claros
- Establecer una relación entre los requerimientos de información ejecutiva y el desempeño del personal de sistemas

7.1.3 Principales problemas

En la implementación de todo proyecto, ya sea de sistemas o no, se presentan problemas, gracias el estudio de campo realizado, se puede identificar que los principales problemas que se enfrentan en la implantación de un sistema de *data warehouse* son:

- Incumplimiento de las fechas establecidas, este es un problema muy común en el área de sistemas por la falta de conceptualización y medición del trabajo intelectual
- Los directivos no se involucran por considerar que los proyectos que tienen que ver con computadoras y *software* no les conciernen



- La resistencia natural al cambio por parte de los usuarios
- La dificultad para integrar y depurar la información que se almacenará en el *data warehouse*

Cabe mencionar que, como en todos los proyectos de desarrollo e implementación de sistemas, debe realizarse un buen análisis de los requerimientos por parte de los diferentes usuarios que tendrán relación con la herramienta, ya que cualquier cambio por parte de estos en una fase avanzada del proyecto puede ocasionar que se deba reestructurar todo lo desarrollado lo cual incrementaría de gran manera el costo y tiempo estimado para la realización del mismo.

7.1.4 Motivos para la implementación

Una empresa grande puede sentir que con la información que cuenta actualmente es suficiente, sin embargo, un buen director de sistemas o un directivo cuya principal actividad es la toma de decisiones, debe saber identificar algunos síntomas que indican claramente que se requiere de "algo más" para realizar el trabajo más eficientemente. Se puede concluir que los principales motivos para la implantación de un sistema de *data warehouse* son:

- Necesidad de la administración efectiva de la información del negocio
- Necesidad de agilización de información para la toma de decisiones
- Gran cantidad de información que, organizada de manera adecuada, puede crear una ventaja competitiva para los directivos
- Necesidad de integrar y centralizar la información de todas las diferentes áreas o unidades del negocio para evitar redundancia e incongruencia de datos
- Necesidad de identificar patrones de ocurrencia o comportamiento que los datos a simple vista no permiten relacionar.

7.2 Recomendaciones

En base a los resultados de la investigación de campo, creo que es pertinente recomendar una autoevaluación empresarial en aquellas empresas que se consideren de clase mundial, o al menos grandes, para identificar el nivel de información con que se cuenta y la cantidad de información que están generando los sistemas transaccionales de manera que podamos calificar si la toma de decisiones actual se realiza en base a información suficiente y relacionada con todo el entorno de la empresa y no solo con algunos aspectos que pudieran no ser base suficiente.

Es un hecho que el avance tecnológico con que cuenta el país no es, por mucho, comparable con el que cuentan algunos países desarrollados, por lo que con la globalización del mercado y la apertura de las fronteras, las empresas deben estar habilitadas para competir con las mejores del mundo, debido a que ser los mejores de una región actualmente no es suficiente, ya que puede encontrarse a un competidor



tan cerca como una computadora con internet esté disponible; por lo que es importante que las empresas se preparen y fundamenten sus decisiones con más y mejor información.

Una recomendación importante es el no adquirir tecnología de información simplemente porque esta de moda o porque los competidores ya la tienen, hay que hacer un estudio minucioso de los requerimientos de información de los niveles ejecutivos, así como estudios de factibilidad reales tanto económicos, técnicos y operativos, para poder sustentar la inversión en una herramienta de estas magnitudes. Si contamos con todo lo anterior, podremos establecer fácilmente un estudio costo-beneficio que refleje la utilidad de la herramienta así como estimar un retorno de inversión muy acercado a lo que sería en realidad.

7.3 Trabajos futuros

Este trabajo de investigación puede servir como fundamento para investigaciones posteriores, ya que está sustentado en lo que la práctica refleja. Algunas de las líneas de investigación que pudieran desarrollarse en base a esta tesis son:

- Realizar un estudio comparativo con una investigación cuantitativa del mismo tipo, pero desarrollada en otra ciudad del país con magnitudes y población similar o con cualquier ciudad del mundo con diferencias culturales, organizacionales y nivel de desarrollo para poder ubicar el estado actual de las empresas de la entidad.
- La realización de una investigación que aplique una encuesta similar a una muestra de la misma entidad, pero con giros de empresas diferentes a las encuestadas en esta investigación , para evaluar las mismas variables y hacer más robusto el estudio, identificando así las tendencias y estados actuales de otro segmento de la población total.
- La creación de una guía para identificar empresas grandes que requieren un nivel de toma de decisiones más preciso, basado en información transaccional, exponiendo claramente la utilidad y bondades que ofrece el contar con una herramienta de *data warehouse*. Esta guía sería usada tanto por los directivos de sistemas como por los encargados de las tomas de decisiones de las empresas para justificar la necesidad de un repositorio de datos centralizado.

Considero que este trabajo cumplió con la finalidad para la que fue creado, y puede ser utilizado confiablemente como referencia para cualquier investigación relacionada con las variables involucradas en la investigación.



Anexo 1

INSTRUMENTO DE INVESTIGACIÓN (ENCUESTA)

Formato de la encuesta

Información general

1. **Numero de empleados de la compañía**
 - a) 500-700
 - b) 700-1,000
 - c) 1,000-5,000
 - d) Más de 5,000

2. **Ventas anuales de la compañía en dólares**
 - a) \$0-\$2,000,000
 - b) \$2,000,000-\$4,000,000
 - c) \$4,000,000-\$10,000,000
 - c) Más de \$10,000,000

3. **¿Cuenta su organización con alguna Tecnología de Información?**
 - a) ERP –Enterprise Resource Planning
 - b) CRM Customer Relationship Management
 - c) SCM - Supply Chain Management
 - d) Ninguno

4. **¿Dónde ubicaría el nivel de Tecnologías de Información con que cuenta su empresa?**
 - a) Transaccional
 - c) Administrativo
 - d) Estratégico

5. **¿Cuenta la empresa con alguna herramienta de *Data Warehouse*? (en caso negativo pasar a pregunta 32)**
 - a) Si
 - b) No

Nivel de implementación

6. **¿Considera importante y/o necesario para su empresa la implementación de un *Data Warehouse*?**
 - a) Si
 - b) No

7. **¿Qué área propuso la implementación del DWH?**
 - a) Sistemas
 - b) Mercadotecnia
 - c) Dirección
 - d) Ventas
 - d) Recursos Humanos
 - e) Producción
 - f) Finanzas
 - g) Otra _____

**8. ¿En qué fase de implantación se encuentra esta herramienta?**

- | | |
|-----------------|-------------------|
| a) Planeación | b) Requerimientos |
| c) Análisis | d) Diseño |
| e) Desarrollo | f) Prueba |
| g) Construcción | h) Otra _____ |

9. El software de su Data Warehouse es:

- | | |
|-------------------------------------|-----------------------|
| a) Paquete comercial | b) Desarrollo interno |
| c) Desarrollo externo (consultoría) | d) Otro _____ |

10. ¿Qué áreas de la empresa están contempladas en el data warehouse?

- | | |
|---|---------------------------------------|
| <input type="checkbox"/> Sistemas | <input type="checkbox"/> Contabilidad |
| <input type="checkbox"/> Ventas | <input type="checkbox"/> Compras |
| <input type="checkbox"/> Recursos humanos | <input type="checkbox"/> Finanzas |
| <input type="checkbox"/> Todas | <input type="checkbox"/> Otra _____ |

11. ¿Qué áreas de la empresa se benefician más con el DWH?

- | | |
|---|---------------------------------------|
| <input type="checkbox"/> Sistemas | <input type="checkbox"/> Contabilidad |
| <input type="checkbox"/> Ventas | <input type="checkbox"/> Compras |
| <input type="checkbox"/> Recursos humanos | <input type="checkbox"/> Finanzas |
| <input type="checkbox"/> Todas | <input type="checkbox"/> Otra _____ |

12. ¿Qué herramienta de explotación de datos ocupan?

- | | |
|--------------------------------------|-------------------------------------|
| <input type="checkbox"/> Data mining | <input type="checkbox"/> OLAP |
| <input type="checkbox"/> ROLAP | <input type="checkbox"/> MOLAP |
| <input type="checkbox"/> Webhousing | <input type="checkbox"/> Otra _____ |

13. ¿Cuánto tiempo duró la implementación del DWH?

- | | |
|---------------------|--------------------------|
| a) Menos de 6 meses | b) Entre 6 meses y 1 año |
| c) De 1 a 3 años | d) De 3 a 5 años |
| e) Más de 5 años | f) No ha terminado |

Motivos para la implementación**14. ¿Qué elementos o variables cree que son necesarios identificar en una empresa para considerar la implementación de un DWH?**

- Administración actual de los sistemas transaccionales.
- Presupuesto para invertir en el proyecto
- Personal especializado en la gestión de *data warehouse*
- Estrategias de negocio orientadas a inversión en tecnologías de almacenamiento masivo de datos
- Todas
- Otra _____



15. ¿Cómo calificaría la productividad generada por el *Data Warehouse*?

Nada 1 2 3 4 5 6 7 8 9 10 Mucha

16. ¿Cree poder detectar una relación *costo-beneficio* al implementar DWH?

Pésima Mala Regular Buena Excelente

17. ¿El *Data Warehouse* implementado, ha mejorado de alguna manera la toma de decisiones del negocio?

a) Si b) No

18. ¿Analizaron diferentes proveedores de DWH antes de adquirir el que tienen?

a) Si b) No

19. ¿Se realizaron estudios de factibilidad?

() Económica () Técnica () Operativa

20. ¿Cuál es el tiempo estimado para el retorno de inversión?

- a) Menos de 6 meses b) Entre 6 meses y 1 año
c) De 1 a 3 años d) De 3 a 5 años
e) Más de 5 años f) No está definido

21. ¿Qué factores hicieron que se decidieran por el que tienen?

- () Precio () Experiencias anteriores
() Recomendación de Zeros () Requerimientos de hardware
() Poder de la herramienta () Otro _____

22. ¿Cuál considera que fue el principal motivo para la implementación del DWH?

- () Centralización de información en un repositorio único de los sistemas transaccionales
() Depuración de la información transaccional
() Necesidad de análisis y toma de decisiones sobre grandes volúmenes de información
() Se presentó la oportunidad en términos de costo de herramientas y consultoría.



Factores críticos de éxito en la implementación

23. ¿Cuáles variables o elementos clave considera valiosos para poder realizar una implementación exitosa de DWH?

- Planeación y administración del proyecto manejada por fases, tomando las variables del negocio como unidades independientes (*Data Mart*)
- Planeación y administración del proyecto manejada por fases, tomando las variables del negocio como una sola unidad (*Data Warehouse*)
- Definición clara de donde se encuentra la información que almacenará el warehouse, las reglas de negocio y los indicadores de desempeño.
- Diseño del repositorio de datos que permite explotar la información de manera multidimensional para alcanzar las capacidades de análisis esperadas.
- Personal con conocimientos de uso y manejo de arquitectura de *data warehouse*.

24. En caso de considerar exitoso su *data warehouse*, ¿cuál cree que fueron los principales motivos de este éxito?

- Alineación entre los requerimientos de información ejecutiva y el trabajo del personal de sistemas.
- Efectiva extracción, transformación y carga de los sistemas transaccionales.
- Herramientas tecnológicas adecuadas que soporten la infraestructura planteada y los volúmenes de información a manejar.
- Coordinación entre el diferente personal involucrado en el proyecto.
- Cultura de uso del *data warehouse* por parte de los gerentes y directores.

25. ¿Quién propició este éxito?

- Personal técnico.
- Líderes de proyecto.
- Gerente de sistemas.
- Usuarios del DW (gerentes y directores).
- Consultores externos de DW

26. ¿Cuál sería un monto aproximado en la inversión de DWH?

- a) Menor de 20,000 us dls.
- b) De 20,000 us a 50,000 us dls.
- c) De 50,000 us a 100,000 us dls.
- d) Más de 100,000 us dls.



Problemas más comunes

27. ¿Enfrentaron problemas en la implementación del DWH? Mencione brevemente algunos. (en caso negativo pasar a pregunta 28)

- a) Si b) No

En caso de ser afirmativo, ¿Cuáles?

- () Resistencia a la implementación por parte del personal o usuarios
- () Falta de involucramiento por parte de los directivos
- () El hardware no cumplió con las necesidades del software
- () No se cumplían las fechas estipuladas
- () Problemas relacionados con la falta de capacitación o experiencia por parte de los implantadores
- () Otro _____

28. ¿Estos problemas son privativos de su empresa o considera que pueden pasarle a cualquier tipo de empresa?

- a) Privativos b) Generales

29. ¿Cuáles considera que son los problemas más importantes o que más obstaculizan la implementación de un DWH?

- | | |
|---------------|-------------|
| 30. Dinero | b) Usuarios |
| c) Dirección | d) Tiempo |
| e) Hardware | f) Software |
| g) Otro _____ | |

30. Por haber participado y compartido su información con la investigación, es posible enviarle los resultados de las demás empresas que contribuyan. ¿Desea que se le haga llegar esa información?

En caso de responder negativamente, pase al inciso 33

31. Datos de la empresa

- Nombre:
- Dirección:
- Teléfono:
- Dirección en Internet :



32. Datos del entrevistado

- Nombre :
- Puesto:
- Función:
- Tiempo desempeñando el cargo
- Personas a su cargo
- Correo electrónico:

33. *Agradezco mucho su tiempo e información. En caso de haberlo requerido, la información se le hará llegar en cuanto se entreviste a la muestra especificada, de lo contrario, le aseguro que su información será tratada solo para fines académicos.*



Referencias utilizadas

- [Adiaans, 1997] • Adiaans, Pieter. DATA MINING 1997
- [Anahory, Sam & Murray, 1997] • Anahory, Sam & Murray, Dennis. Data warehousing in the real world: a practical guide for building decision support systems. Addison Wesley, 1997
- [Anthem Inc, 2001] • Anthem Inc., BUSINESS LINE: India: Scoop up data, good health Businessline, Islamabad , Mar 7, 2001
- [ATG, 1998] • ATG 1998, "Managing the Warehouse thru its Lifecycle" ATG's Data Warehousing Technology Guide. 1998
- [ATG, 1999] • ATG 1999, "Building a Decision Support Architecture for Data Warehousing" ATG's Data Warehousing Technology Guide. 1999
- [ATG, 2000] • ATG 2000, "Enterprise Storage for Today's Data Warehousing Environment" ATG's Data Warehousing Technology Guide. 2000
- [Computerworld, 1999] • Computerworld 1999, "Informe Data Mining" Suplemento 709 ComputerWorld España (1999).
- [Date, 1990] • Date C. J.. Database: A premier. Addison Wesley. 1990
- [Date, 1986] • Date C. J.. Introducción a los sistemas de bases de datos. Addison Wesley. 1986
- [Deakin, 1998] • Deakin, Rose. Bases de datos, México 1998
- [Debevoise, 1999] • Debevoise, Tom. The data warehouse method : integrated data warehouse support environments. Upper Saddle River, N.J. : Prentice Hall PTR, c1999
- [Devlin, 2000] • Devlin, Barry. Data warehouse : from architecture to implementation Reading, Mass : Addison-Wesley, 2000



-
- [Fernandez, 1999] • Jorge C. Fernández, "El Impacto de Business Intelligence en el Negocio", Revista Soluciones avanzadas, México, D.F., 15- Junio-1999
- [Fortier, 1997] • Fortier, Paul. Database systems handbook. Mc Graw Hill. 1997
- [Gray, 1998] • Gray, Paul, Decision support in the data warehouse Upper Saddle River, NJ : Prentice Hall, 1998
- [Groth, 2000] • Groth, Robert. Data mining : building competitive advantage Upper Saddle River, NJ : Prentice Hall PTR, 2000
- [Inmon, 1996] • Inmon, William H . Building the data warehouse New York : Wiley Computer Pub., 1996, 2nd ed
- [Inmon, 1997] • W.H. Inmon, J.D. Welch, Katherine L. Glassey, New York : Wiley, 1997
- [Inmon, 1999] • Inmon. Data warehouse performance New York : John Wiley, [1998], c1999
- [Inmon, 2000] • Inmon, William H. Managing the data warehouse : New York : John Wiley, 2000
- [ITESO, 1998] • ITESO. 1998.
<http://www.gdl.iteso.mx/biblio/bases.html>
- [Kimball, 1996] • Kimball, Ralph . The data warehouse toolkit : practical techniques for building dimensional data warehouses. New York : John Wiley & Sons, 1996
- [Kimball, 1998] • Kimball, Ralph... [et al.] The data warehouse lifecycle toolkit : expert methods for designing, developing, and deploying data warehouses New York : Wiley, c1998
- [LGI, 1998] • LGI 1998 "Data Warehousing Information Center" LGI System Incorporated, 1998
- [Marcos, 1998] • Marcos, María del Socorro. Manual para la elaboración de Tesis, Editorial Trillas, México, D.F., 1998.



-
- [Mattison,1999]
- Mattison, Rob. Web warehousing and knowledge management / Rob Mattison, New York : McGraw-Hill, 1999
- [Microsoft, 1998]
- Microsoft. The MS Data Warehousing Strategy : A platform for improved decision-making through easier data access and analysis. 1998.
<http://www.microsoft.com/TechNet/sql/Technote/datawhst.asp>
- [Microstrategy, 1999]
- "The Case For Relational OLAP" MicroStrategy Incorporated White Paper. 1999
- [Norsistemas, 1998]
- NorSistemas 1998, "Data Mining" Curso Formación Interno NorSistemas. 1998
- [Poe, 1998]
- Poe, Vidette. Building a data warehouse for decision support, Upper Saddle River, NJ Prentice Hall PTR, c1998, 2nd ed
- [SAS, 1998]
- SAS 1998, "Rapid Data Warehousing with the SAS System" SAS Institute (1998). [En línea]
<http://www.sas.com/service/library/periodicals/sascomm/1q98/rapid.html>
- [NCR, 2000]
- NCR, Customer Success Stories: Office Depot [En línea] :
<http://www.teradatalibrary.com/pdf/eb3030.pdf>, 2000, (Consultado: Diciembre, 2001)
- [NCR, 2001]
- NCR, Customer Success Stories: Delta Air Lines [En línea]:
<http://www.teradatalibrary.com/pdf/eb1310.pdf>, 2000 (Consultado: Diciembre, 2001)