



UNIVERSIDAD DEL SALVADOR
FACULTAD DE CIENCIAS ECONÓMICAS

MAESTRÍA EN AUDITORÍA DE SISTEMAS

TESIS DE MAESTRÍA

TEMA:

***“DATA WAREHOUSE: RIESGOS
INHERENTES DE UNA IMPLEMENTACIÓN
Y CONTROLES PARA MITIGARLOS”***

UNIVERSIDAD
DEL SALVADOR

AUTORA:

ING. STELLA MARIS PARENTI

TUTOR:

DR. GUSTAVO FRANCISCO MANNISE

ABRIL / 2007



INDICE

1 INTRODUCCIÓN.	3
2 OBJETIVO DE LA TESIS.	6
3 CONCEPTO DE DATA WAREHOUSE.	7
3.1 Análisis de las principales características del DW.	8
3.1.1 Integrado.	8
3.1.2 Temático.	10
3.1.3 Histórico o de Tiempo Variante.	11
3.1.4 No Volátil.	13
3.2 Objetivos del DW.	14
3.2.1 Objetivos a Corto Plazo.	14
3.2.2 Objetivos a Largo Plazo.	16
3.3 Justificaciones para el uso de un DW.	17
3.4 El DW como sistema de misión crítica.	17
4 USOS Y APLICACIONES DE UN DATA WAREHOUSE.	18
4.1 Sólo pocos usuarios acceden a los datos concurrentemente.	18
4.2 Los usuarios generan un procesamiento no predecible complejo.	18
4.3 Las consultas de los usuarios acceden a cantidades grandes de datos.	19
4.4 Las consultas de los usuarios no tienen tiempos de respuesta críticos.	19
5 CONVERSIÓN DE DATOS EN INFORMACIÓN.	21
5.1 La evolución del soporte de decisión.	21
5.2 Técnicas para utilizar un DW.	24
5.3 Características de los datos.	25
5.4 Fuentes de datos para análisis.	26
5.5 Lo que cuesta alcanzar la cima.	27
6 EL VALOR DE LA INFORMACIÓN PARA LA TOMA DE DECISIONES.	27
6.1 DW, soporte de decisión e inteligencia comercial.	27
7 OLTP VS. OLAP.	28
7.1 Definiciones.	28
7.2 Diferencias.	29
7.3 Relación entre ambos esquemas.	32
7.4 Ambientes compartidos o no.	32
8 LAS BASES PARA EL DESARROLLO DE UN DATA WAREHOUSE.	33
8.1 Estructura del DW.	33
8.2 Arquitectura del DW.	36
8.3 Construcción de un DW.	40
8.3.1 Ciclo de Vida.	40
8.4 Metodología para construir un DW.	49
8.4.1 Esquemas.	54
8.4.2 Opciones de implementación.	56
8.4.3 La alimentación y el acceso.	56
8.4.4 Las infraestructuras.	62
8.4.5 Mantenimiento del DW.	63
8.5 Funciones durante el desarrollo.	63
8.5.1 Funciones de Gerencia.	64
8.5.2 Funciones de Elaboración de Requerimientos.	64
8.5.3 Funciones de Desarrollo de Base de Datos.	65
8.5.4 Funciones de Desarrollo de la Aplicación.	65



8.6 Consideraciones adicionales para construir y administrar un DW.....	66
9 HERRAMIENTAS PARA LA PRE Y POST- IMPLEMENTACIÓN.....	67
9.1 Herramientas de modelado y diseño de datos.....	67
9.2 Herramientas de extracción y carga de datos.....	68
9.3 Herramientas de gestión y administración.....	69
10 LA IMPORTANCIA DE LA ADMINISTRACIÓN DE METADATOS.....	71
10.1 Por qué son importantes.....	72
10.2 Almacenamiento y administración persistente de los metadatos.....	73
10.3 Estándares de metadatos.....	74
11 COSTOS VERSUS VALOR DEL DATA WAREHOUSE.....	75
11.1 Costos de Construcción.....	75
11.2 Costos de Mantenimiento y Operación.....	76
11.3 Valor del DW.....	77
11.4 Balance de Costos versus Valor.....	79
12 IMPACTOS DE IMPLEMENTACIÓN DEL DATA WAREHOUSE.....	79
12.1 Humanos.....	80
12.2 Empresariales.....	80
12.3 Comunicación e Impactos Organizacionales.....	81
12.4 Impactos Técnicos.....	81
13 EL DATA MINING Y SU RELACIÓN CON EL DATA WAREHOUSE.....	82
14 AUDITORIA DEL DATA WAREHOUSE.....	83
14.1 Guía de riesgos y mejores prácticas.....	84
14.1.1 Riesgos.....	84
14.1.2 Mejores prácticas.....	85
14.2 Descripción de riesgos y controles para las distintas etapas del Ciclo de Vida del DW (Pre-Implementación).....	85
14.3 Descripción de riesgos y controles durante la explotación del DW (Post-Implementación).....	112
14.4 Técnicas para realizar la auditoria.....	128
15 CONCLUSIONES.....	129
16 GLOSARIO.....	134
17 BIBLIOGRAFIA.....	148



“DATA WAREHOUSE: RIESGOS INHERENTES DE UNA IMPLEMENTACIÓN Y CONTROLES PARA MITIGARLOS”

1 INTRODUCCIÓN.

Como consecuencia de la gran cantidad de datos y la creciente demanda de sistemas que generen información de mejor calidad, las empresas requieren el diseño de estructuras de datos y su integración a través de distintos repositorios de manera eficiente, con el fin de tener acceso a una visión histórica y sobre distintos aspectos o dimensiones del negocio con el objetivo de crear indicadores de gestión y alimentar de información a modelos para mejorar los procesos de marketing, rentabilidad, control de riesgos y contribuir a la toma de decisiones estratégicas, basadas en datos íntegros y coherentes (o consistentes). Como consecuencia de la necesidad de esta integración nace el Data Warehouse y el Data Warehousing.

Un Data Warehouse (DW) es un depósito de datos e información transaccional, operativa y externa, orientado a un dominio, integrado, no volátil y variable en el tiempo, que se encuentra disponible para consultas y análisis para la toma de decisiones de la empresa u organización, mientras que el Data Warehousing es el proceso que recopila datos de varias aplicaciones, en los sistemas operativos de una organización, integra la información en un modelo lógico de áreas de tópicos comerciales, la almacena de modo que resulte accesible a los tomadores de decisiones y se los provee a través de herramientas de consulta y generación de informes (en la presente tesis se hablará de Data Warehouse y de Data Warehousing indistintamente). Los datos y la información se extraen de fuentes heterogéneas y se almacenan en este único repositorio de forma tal que sea más fácil y más eficiente realizar consultas sobre los datos.

La creación de un DW se fundamenta en dos ideas:

- Integración de los datos de bases de datos distribuidas, heterogéneas y diferentemente estructuradas, que facilita una descripción y análisis global.
- Separación de los datos usados en operaciones diarias de los datos usados en el DW para análisis y toma de decisiones corporativas.

Hacer consultas a las bases de datos de los sistemas operativos puede afectar y empeorar



el tiempo de respuesta de estos sistemas críticos de las empresas y pueden tener limitaciones en cuanto a la cantidad de datos históricos disponibles.

El diseñar un DW no constituye una tarea sencilla; combinar datos de distintas fuentes suele ser bastante complicado para las personas encargadas de tomar decisiones. Normalmente hay que homogeneizar los datos de una forma u otra y en el DW esto se hace durante el proceso de carga. Como consecuencia de estas estrategias se tarda mucho menos en acceder a los datos del DW que en hacer una consulta a varias bases de datos distintas.

Periódicamente se importan datos de los sistemas transaccionales y de otras fuentes de datos relacionadas al negocio, en el DW, mediante el proceso de Extracción, Transformación y Carga o ETL (por las siglas en inglés de Extraction, Transformation and Loading), que transforma los datos del modelo operativo al modelo de análisis. Es una práctica común realizar un almacenamiento temporario o Data Storage con los datos extractados para poder transformarlos, limpiarlos y procesarlos antes de combinarlos en el DW.

Debido a que las empresas no siempre cuentan con aplicaciones únicas de gestión, pueden poseer replicaciones y distintos sistemas para atender distintos conjuntos de operaciones, corriéndose el riesgo que las bases de datos de los sistemas operacionales contengan datos duplicados, a veces erróneos, superfluos o incompletos, más allá de los riesgos propios de cualquier interfase o transformación. En ocasiones los riesgos mencionados hacen necesario aplicar controles para realizar el ajuste de los datos para posibles comparaciones, por ejemplo distintas monedas o sistemas contables.

Los datos históricos no suelen guardarse en los sistemas operacionales, pero son elemento esencial para cualquier análisis. El DW es el lugar adecuado para estos datos, donde se guardan además de los datos de detalle, los agregados que son necesarios para minimizar su almacenamiento, optimizar el tiempo de respuesta al usuario, así como la comprensión de sus datos.

Otro proceso que se aplica es el de limpieza de datos el cual es un proceso multidisciplinario y complejo. En este proceso se deben analizar los datos para evitar o



disminuir los riesgos de inexactitudes, anomalías y otros problemas para luego transformar esos datos y asegurar que sean precisos y coherentes. Asegurar la integridad referencial, validar los datos, producir la meta data o descripción de los datos, formato y significado relacionado al negocio son sólo algunas de las complejas tareas a realizar y controlar.

Entre los controles a aplicar cuando se realiza el proceso de integración de datos desde distintas fuentes, se encuentra la integración de los diferentes esquemas para asegurar que entidades equivalentes se aparean correctamente cuando se produce la fusión desde distintas fuentes de datos. En ocasiones también es necesario determinar la redundancia mediante correlaciones. Un adecuado control y atención en el proceso de integración de datos reducirá o evitará redundancias e inconsistencias en los análisis resultantes a partir de estos datos.

A diferencia del desarrollo de aplicaciones, donde los requerimientos de la empresa logran ser relativamente bien definidos producto de la estabilidad de las reglas de negocio a través del tiempo, construir un DW depende tanto de la realidad de la empresa como de las condiciones que en ese momento existan, las cuales determinan qué debe contener el DW; por lo tanto su construcción requiere la participación activa de quienes lo utilizarán, con el fin que responda a la emisión de informes y generación de información útil para la toma de decisiones y evitar el riesgo de pérdida de tiempo en la espera de información que finalmente es incorrecta o no encontrada. El analista debe saber interpretar los requerimientos del usuario de cómo van a ser usados los datos, qué tipo de consultas se realizarán, qué significa la relación entre un dato y otro, etc. con el objetivo de realizar el modelo y la estructura que sustente la base que permita al analista construirla y a los ejecutivos consultarla. Debe tenerse presente que la información contenida en el DW puede ser requerida en cualquier momento. Si bien en el ambiente de DW las peticiones de los usuarios no tienen un requerimiento de respuesta crítico, porque el resultado frecuentemente se usa en un proceso de análisis y toma de decisiones, éstos esperan una respuesta dentro del mismo día en que es hecha la consulta. Además puede suceder que una transacción del ambiente operacional sea ingresada al DW con un retraso por ejemplo de veinticuatro horas, con lo cual los nuevos datos no estarían disponibles en el DW durante ese período. Este hecho trae



consigo, si no se aplican los controles correspondientes, el riesgo que la información no sea oportuna, es decir que no se encuentre disponible en tiempo y forma en el momento en que el nivel decisorio de la organización la necesite.

Por otra parte, las organizaciones empresariales y la gente de la cual ella se compone queda determinada por el acceso a la información. De esta manera, la gente queda mejor habilitada para entender su propio rol y responsabilidades como también los efectos de sus contribuciones; a la vez que desarrollan un mejor entendimiento y apreciación con las contribuciones de otros.

Dado que el DW reúne datos de distinta sensibilidad y teniendo en cuenta los diferentes roles y responsabilidades mencionados en el párrafo anterior, se debe considerar el riesgo de acceso no autorizado a dichos datos o a la información que ellos proveen. Para mitigar el riesgo de acceso no autorizado se debería aplicar una combinación de control a nivel de acceso físico y a nivel de aplicación. Las decisiones sobre el control de acceso están basadas en las políticas organizacionales, en la segregación de funciones y el mínimo privilegio. Por otro lado, debe tenerse en cuenta que para que los controles sean efectivos no deben interrumpir el normal desempeño de los usuarios más de lo necesario y no debe ser una carga ni para los administradores, auditores o usuarios autorizados.

Algunos otros riesgos inherentes que pueden presentarse a lo largo de las distintas etapas del Ciclo de Vida del DW, sumados a los mencionados precedentemente, son la pérdida de valores, existencia de datos extraños fuera de formato o de rango, etc., para lo cual es necesario aplicar controles que mitiguen aquellos riesgos con el objeto de garantizar la integridad, exactitud, oportunidad y el acceso a la información que será utilizada por los distintos niveles jerárquicos de una organización, en la toma de decisiones corporativas y de estrategias de negocio.

2 OBJETIVO DE LA TESIS.

La presente tesis tiene como objetivo dar un panorama general y actualizado de los componentes de un DW y responder a los siguientes interrogantes:



1. ¿Cuán importante es la participación del usuario final y la unificación de criterios en la definición de las funcionalidades que debe proveer el DW?
2. ¿En qué medida la aplicación de controles adecuados en cada una de las etapas del Ciclo de Vida del DW, puede contribuir a minimizar los riesgos inherentes (integridad, exactitud, oportunidad, acceso) que se presentan en cada una de las mismas, para garantizar la calidad de los datos almacenados y utilizados para la toma de decisiones?
3. ¿De qué forma se mitigan los riesgos (integridad, exactitud, oportunidad, acceso) del tratamiento de datos con el fin de convertirlos en información útil para agregarle valor al negocio y generar el DW?
4. ¿Se justifica el costo de la implementación de un DW?

3 CONCEPTO DE DATA WAREHOUSE.

El DW soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico, informático sobre una amplia perspectiva de tiempo.

Existen varias definiciones para un DW, la más conocida es la propuesta por quien es considerado el padre de las Bases de Datos:

"Un DW es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales" (Inmon, 1992).

Por otro lado tenemos una definición que refleja claramente el principal beneficio que el DW aporta a la empresa, eliminar aquellos datos que obstaculizan la labor de análisis de información y entregar la información que se requiere en la forma más apropiada, facilitando así el proceso de gestión.

Esta es, *"Considerar al DW como algo que provee dos beneficios empresariales reales: Integración y Acceso de datos. DW elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operativo clásico"* (Osterfeldt, 1993).



Las definiciones anteriores, deben tomarse como definición pura de DW, pero algunos términos son manejados según las necesidades y capacidades del mercado dando origen al concepto de Datamart, el cual representa un conjunto de hechos y datos organizados para dar soporte a las decisiones. Estos datos son orientados a satisfacer las necesidades particulares de un proceso de negocio específico. Es un subconjunto lógico del DW.

Un DW se caracteriza por:

- Ser Integrado - Temático - Histórico - No volátil.
- No ser un producto, sino un proceso para consolidar y administrar datos de variadas fuentes con el propósito de responder preguntas de negocios y tomar decisiones.
- Consolidar datos desde una variedad de fuentes.
- Manejar grandes volúmenes de datos de una forma que no era posible, o era costoso. A estos medios los agruparemos en Procesamiento y Administración de Datos.
- Acceder a los datos de una forma más directa, en "el lenguaje del negocio" y analizarlos para obtener relaciones complejas entre los mismos.

3.1 Análisis de las principales características del DW.

3.1.1 Integrado.

Los datos almacenados en el DW deben integrarse en una estructura consistente, las inconsistencias existentes entre los diversos sistemas operativos deben ser eliminadas.

El aspecto más importante del ambiente DW es que la información encontrada en el interior está siempre integrada. Esta integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros. A través de los años, los diseñadores de las diferentes aplicaciones han tomado sus propias decisiones sobre cómo se debería construir una aplicación. Los estilos y diseños personalizados se muestran de muchas



maneras.

Se diferencian en la codificación, en las estructuras claves, en sus características físicas, en las convenciones de nombramiento y otros.

Codificación. Los diseñadores de aplicaciones codifican el campo GENERO en varias formas. Un diseñador representa GENERO como una "M" y una "F", otros como un "1" y un "0", otros como una "X" y una "Y" e inclusive, como "masculino" y "femenino".

No importa mucho cómo el GENERO llega al DW. Probablemente "M" y "F" sean tan buenas como cualquier otra representación. Lo importante es que cualquiera sea la fuente de donde venga, el GENERO debe llegar al DW en un estado integrado uniforme. Por lo tanto, cuando el GENERO se carga en el DW desde una aplicación, donde ha sido representado en formato "M" y "F", los datos deben convertirse al formato del DW.

Medida de atributos. Los diseñadores de aplicaciones miden las unidades de medida en una variedad de formas. Un diseñador almacena los datos en centímetros, otros en pulgadas, otros en millones de pies cúbicos por segundo y otros en yardas.

Al dar medidas a los atributos, la transformación traduce las diversas unidades de medida usadas en las diferentes bases de datos en una medida estándar común.

Cualquiera sea la fuente, cuando la información llegue al DW necesitará ser medida de la misma manera.

Convenciones de Nombramiento. El mismo elemento es frecuentemente referido por nombres diferentes en las diversas aplicaciones. El proceso de transformación asegura que se use preferentemente el nombre de usuario.

Fuentes Múltiples. El mismo elemento puede derivarse desde fuentes múltiples. En este caso, el proceso de transformación debe asegurar que sea usada la fuente apropiada, documentada y movida al depósito.

Los puntos de integración afectan casi todos los aspectos de diseño, las características físicas de los datos, la disyuntiva de tener más de una fuente de datos, el problema de



estándares de denominación inconsistentes, formatos de fecha inconsistentes y otros. Cualquiera sea la forma del diseño, el resultado es el mismo, la información necesita ser almacenada en el DW en un modelo globalmente aceptable y singular, aun cuando los sistemas operacionales subyacentes almacenen los datos de manera diferente.

Cuando el analista del sistema de soporte de decisiones observe el DW, su enfoque deberá estar en el uso de los datos que se encuentren en el depósito, antes que preguntarse sobre la confiabilidad o consistencia de los datos.

3.1.2 Temático.

Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales y en base a los aspectos que son de interés para la empresa. Siendo así, los datos tomados están en contraste con los clásicos procesos orientados a las aplicaciones. En la figura se muestra el contraste entre los dos tipos de orientaciones.



El ambiente operativo se diseña alrededor de las aplicaciones y funciones tales como préstamos, ahorros, tarjeta bancaria y depósitos para una institución financiera. Por



ejemplo, una aplicación de ingreso de órdenes puede acceder a los datos sobre clientes, productos y cuentas. La base de datos combina estos elementos en una estructura que acomoda las necesidades de la aplicación.

En el ambiente DW se organiza alrededor de sujetos tales como cliente, vendedor, producto y actividad. Por ejemplo, para un fabricante, éstos pueden ser clientes, productos, proveedores y vendedores. Para una universidad pueden ser estudiantes, clases y profesores. Para un hospital pueden ser pacientes, personal médico, medicamentos, etc. La alineación alrededor de las áreas de los temas afecta el diseño y la implementación de los datos encontrados en el DW. Las principales áreas de los temas influyen en la parte más importante de la estructura clave.

Las aplicaciones están relacionadas con el diseño de la base de datos y del proceso. En DW se enfoca el modelado de datos y el diseño de la base de datos. El diseño del proceso (en su forma clásica) no es separado de este ambiente.

Las diferencias entre la orientación de procesos y funciones de las aplicaciones y la orientación a temas, radican en el contenido de la data a nivel detallado. En el DW se excluye la información que no será usada por el proceso de sistemas de soporte de decisiones, mientras que la información de las orientadas a las aplicaciones, contiene datos para satisfacer de inmediato los requerimientos funcionales y de proceso, que pueden ser usados o no por el analista del sistema de soporte de decisiones.

Otra diferencia importante está en la interrelación de la información. Los datos operativos mantienen una relación continua entre dos o más tablas basadas en una regla comercial que está vigente. Las del DW miden un espectro de tiempo y las relaciones encontradas en el DW son muchas. Muchas de las reglas comerciales (y sus correspondientes relaciones de datos) se representan en el DW, entre dos o más tablas.

3.1.3 Histórico o de Tiempo Variante.

El tiempo es parte implícita de la información contenida en un DW. En los sistemas operativos, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. La información almacenada en el DW sirve, entre otras cosas, para



realizar análisis de tendencias. El DW se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

Esta característica básica de los datos en un depósito, es muy diferente de la información encontrada en el ambiente operativo. En éstos, la información se requiere al momento de acceder. En otras palabras, en el ambiente operativo, cuando usted accede a una unidad de información, espera que los valores requeridos se obtengan a partir del momento de acceso.

Como la información en el DW es solicitada en cualquier momento (es decir, no "ahora mismo"), los datos encontrados en el depósito se llaman de "tiempo variante".

Los datos históricos son de poco uso en el procesamiento operativo. La información del depósito por el contraste, debe incluir los datos históricos para usarse en la identificación y evaluación de tendencias.

El tiempo variante se muestra de varias maneras:

La más simple es que la **información representa los datos sobre un horizonte largo de tiempo**, desde cinco a diez años. El horizonte de tiempo representado para el ambiente operativo es mucho más corto, desde valores actuales hasta sesenta a noventa días.

Las aplicaciones que tienen un buen rendimiento y están disponibles para el procesamiento de transacciones, deben llevar una cantidad mínima de datos si tienen cualquier grado de flexibilidad. Por ello, las aplicaciones operativas tienen un corto horizonte de tiempo, debido al diseño de aplicaciones rígidas.

La segunda manera en la que se muestra el tiempo variante en el DW está en la **estructura clave** que contiene, implícita o explícitamente, un elemento de tiempo como día, semana, mes, etc. Este elemento de tiempo está casi siempre al pie de la clave concatenada, encontrada en el DW. En ocasiones, existirá implícitamente, como el caso en que un archivo completo se duplica al final del mes o al cuarto.

La tercera manera es cuando **la información del DW, una vez registrada correctamente, no puede ser actualizada**. La información del DW es, para todos los propósitos prácticos, una serie larga de "snapshots" (vistas instantáneas).

Por supuesto, si los snapshots de los datos se han tomado incorrectamente, entonces



pueden ser cambiados. Asumiendo que los snapshots se han tomado adecuadamente, ellos no son alterados una vez hechos. En algunos casos puede ser no ético e incluso ilegal, alterar los snapshots en el DW. Los datos operativos, siendo requeridos a partir del momento de acceso, pueden actualizarse de acuerdo a la necesidad.

3.1.4 No Volátil.

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

Los datos operativos cambian sobre una base de datos momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una base de datos estable.

La actualización (insertar, borrar y modificar), se hace regularmente en el ambiente operativo sobre una base de datos registro por registro. Pero la manipulación básica de los datos que ocurre en el DW es mucho más simple. Hay dos únicos tipos de operaciones: la carga inicial de datos y el acceso a los mismos. No hay actualización de datos (en el sentido general de actualización) en el depósito, como una parte normal de procesamiento.

Hay algunas consecuencias muy importantes de esta diferencia básica, entre el procesamiento operativo y del DW. En el nivel de diseño, la necesidad de ser precavido para actualizar las anomalías no es un factor en el DW, ya que no se hace la actualización de datos. Esto significa que en el nivel físico de diseño, se pueden tomar libertades para optimizar el acceso a los datos, particularmente al usar la normalización y desnormalización física.

Otra consecuencia de la simplicidad de la operación del DW está en la tecnología subyacente, utilizada para correr los datos en el depósito. Teniendo que soportar la actualización de registro por registro en modo on-line se requiere que la tecnología tenga un fundamento muy complejo debajo de una fachada de simplicidad.

La tecnología permite realizar back-up y recuperación, transacciones e integridad de los datos y la detección y solución al estancamiento que es más complejo. En el DW no es



necesario el procesamiento.

La fuente de casi toda la información del DW es el ambiente operativo. A simple vista, se puede pensar que hay redundancia masiva de datos entre los dos ambientes. Dicho razonamiento es superficial. De hecho, hay una mínima redundancia de datos entre ambos ambientes.

Se debe considerar lo siguiente:

- Los datos se filtran cuando pasan desde el ambiente operativo al de depósito. Existe mucha data que nunca sale del ambiente operativo. Sólo los datos que realmente se necesitan ingresarán al ambiente de DW.
- El horizonte de tiempo de los datos es muy diferente de un ambiente al otro. La información en el ambiente operativo es más reciente con respecto a la del DW. Desde la perspectiva de los horizontes de tiempo únicos, hay poca superposición entre los ambientes operativo y de DW.
- El DW contiene un resumen de la información que no se encuentra en el ambiente operativo.
- Los datos experimentan una transformación fundamental cuando pasan al DW. La mayor parte de los datos se alteran significativamente al ser seleccionados y movidos al DW, por lo que no es la misma data que reside en el ambiente operativo desde el punto de vista de integración.

En vista de estos factores, la redundancia de datos entre los dos ambientes es una ocurrencia rara, que resulta en menos de 1%.

3.2 Objetivos del DW.

Los clasificaremos en objetivos a **Corto Plazo** siendo aquellos que se producen con cada iteración del DW, representando un resultado inmediato para el usuario. Y en objetivos a **Largo Plazo** siendo los resultados a los que se arriba mediante el cumplimiento de los otros objetivos.

3.2.1 Objetivos a Corto Plazo.

Mejorar la calidad de la información: Una de las deficiencias de las Bases de Datos