

CARLOS OSWALDO MEDINA GUERRERO

**APLICACION DE UN DATA WAREHOUSE
ORIENTADO A LA BANCA**



universidad
FRANCISCO MARROQUIN

**FACULTAD DE INGENIERIA DE SISTEMAS,
INFORMATICA Y CIENCIAS DE LA COMPUTACIÓN**

GUATEMALA, 2002

Esta tesis fué elaborada por el autor como requisito para obtener el título de Licenciado en Administración de Sistemas de Información.

Guatemala, febrero de 2002



UNIVERSIDAD FRANCISCO MARROQUIN

FACULTAD DE INGENIERIA DE SISTEMAS,
INFORMATICA Y CIENCIAS DE LA COMPUTACION

Guatemala, 11 de Abril de 2000

Señor
Carlos Oswaldo Medina Guerrero
Presente.-

Estimado señor Medina:

Tengo mucho gusto en informarle que ha sido autorizado su punto de Tesis, previo a optar al grado académico de Licenciado en Administración de Sistemas de Información, cuyo título es "Aplicación de un Data Warehouse Orientado a la Banca".

Al mismo tiempo le informo que ha sido aprobada la designación de la Ingeniera María Mercedes Zaghi como asesora de su trabajo de graduación.

Atentamente,

FACULTAD DE INGENIERIA DE SISTEMAS,
INFORMATICA Y CIENCIAS DE LA COMPUTACION

Dr. Eduardo Sager
Decano



Guatemala, 30 de enero del 2002

Dr.
Eduardo Suger Cofiño
Decano de FISICC
Universidad Francisco Marroquín
Presente

Estimado Dr. Suger:

He revisado la parte teórica del trabajo de tesis titulado "Aplicación de un Data Warehouse orientado a la Banca", presentada por Carlos Oswaldo Medina Guerrero, carné No. 1592160. A mi criterio, cumple con los requisitos exigidos por esta facultad.

En mi calidad de asesor, someto a Ud., el presente trabajo para su aprobación.

Atentamente,


Ing. María Mercedes Zaghi G.



UNIVERSIDAD FRANCISCO MARROQUIN

FACULTAD DE INGENIERIA DE SISTEMAS,
INFORMATICA Y CIENCIAS DE LA COMPUTACION

Guatemala, 11 de febrero de 2002.

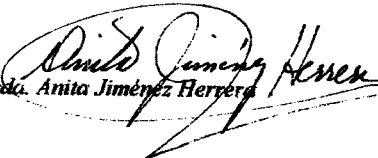
Doctor
Eduardo Suger Cofiño
Decano de la Facultad de Ingeniería
de Sistemas, Informática y Ciencias
de la Computación

Doctor Suger:

Por este medio, me permito comunicarle que leí la tesis del alumno CARLOS OSWALDO MEDINA GUERRERO carné # 1592160 titulada "APLICACIÓN DE UNA DATA WAREHOUSE ORIENTADA A LA BANCA" asesorada por la Ingeniera Maria Mercedes Zaghi G.

Después de revisarla detenidamente y de hacer las recomendaciones pertinentes, en mi calidad de revisora de redacción, estilo y ortografía, le comunico que dicha tesis llena los requisitos que exige la Universidad.

Me suscribo del Señor Decano, como su atenta y segura servidora.


Lidia Anita Jiménez Herrera

AJH/cr.

c.c. Archivo



UNIVERSIDAD FRANCISCO MARROQUIN

FACULTAD DE INGENIERIA DE SISTEMAS,
INFORMATICA Y CIENCIAS DE LA COMPUTACION

Guatemala, 11 de febrero de 2002

Señor
Carlos Oswaldo Medina Guerrero
Presente

Estimado señor Medina:

Tengo mucho gusto en informarle que, después de haber revisado su trabajo de Tesis, cuyo título es **“APLICACION DE UNA DATA WAREHOUSE ORIENTADA A LA BANCA”**, y de haber obtenido el dictamen del asesor específico, autorizo la publicación del mismo.

Aprovecho la oportunidad para felicitarlo por el magnífico trabajo realizado, el cual es de indiscutible beneficio para el desarrollo de las Ciencias de Computación en Guatemala.

Atentamente,

FACULTAD DE INGENIERIA DE SISTEMAS,
INFORMATICA Y CIENCIAS DE LA COMPUTACION



Dr. Eduardo Suger C.
Decano FISICC

ESC/cdec.-

ÍNDICE

CAPITULOS	Página
CAPITULO I	
1. Introducción.....	
1.1 Justificación histórica	
1.2 Sumario.....	3
CAPITULO II	
2. Análisis del entorno Data Warehouse.....	4
2.1 ¿Qué es un Data warehouse?	4
2.1.1 Orientado a sujetos o temas (tematico).....	5
2.1.2 Integrado	8
2.1.2.1 Codificación.....	10
2.1.2.2 Medida de los atributos	10
2.1.3 Variante en el tiempo (historico)	11
2.1.4 No volatil.....	14
2.2 La estructura del Data Warehouse.....	17
2.2.1 Almacenamiento.....	17
2.2.2 Flujo de Datos.....	23
2.2.3 Utilización del Data Warehouse.....	24
2.3 Conceptos y funcionalidades básicas	33
2.3.1 La función bancaria	33
2.3.2 Aspectos a considerar	35
2.3.2.1 Cambio de hábitos en la toma de decisiones.....	35

2.3.2.2 Cambio en la forma de desarrollar soluciones de la tecnología de la información	38
2.3.2.3 Alta integración entre los componentes	39
2.3.3 Formación del equipo y personas involucradas.....	40
2.3.4 Gestión del proyecto.....	49
2.3.5 Data Warehouse vrs. data mart.....	52
2.3.6 Componentes a tener en cuenta a la hora de construir un Data Warehouse orientado a la banca	55
2.3.6.1 Hardware.....	55
2.3.6.2 Software de almacenamiento (SGBD)	57
2.3.6.3 Software de extracción y manipulación de datos	60
2.3.6.4 Herramientas middleware	61
2.4 Construyendo el Data Warehouse orientado a la banca	64
2.4.1 Fases de implantación del Data Warehouse	64
2.4.1.1 Definición de los objetivos.....	66
2.4.1.2 Definición de los requerimientos de información.....	66
2.4.1.3 Diseño y modelado	66
2.4.1.4 Implantación.....	67
2.4.1.5 Revisión	68
2.4.1.6 Diseño de la estructura de cursos de formación	69
2.4.2 Estrategias de implantación.....	69
2.4.3 Técnicas de explotación de la información	70
2.4.3.1 OLAP, ROLAP, MOLAP	71
2.4.3.2 Consultas y reportes	79

2.4.3.3 Data Mining o Minería de Datos.....	80
2.4.3.4 Webhousing	87
2.4.4 Tipos de aplicaciones en las que utilizar técnicas disponibles sobre el Data Warehouse orientado a la banca.....	90
2.4.4.1 Data Warehouse y Sistemas de Mercadeo.....	90
2.4.4.2 Data Warehouse y Análisis de Riesgo Financiero	91
2.4.4.3 Data Warehouse y Análisis de Riesgo de Crédito.....	93
2.4.4.4 Data Warehouse: Otras áreas de aplicación.....	95
2.4.5 Seguridad de acceso y manipulación de la información en el Data Warehouse	96

CAPITULO III

3. Aspectos técnicos en el proceso de adquisición de sistemas de Data Warehouse.....	98
3.1 Análisis de las necesidades del comprador	98
3.1.1 Definición de los objetivos	99
3.1.2 Definición de los requerimientos de información	100
3.2 Factores relevantes en el proceso de adquisición	100
3.2.1 Pruebas en condiciones reales.....	101
3.2.2 Volumen y organización de los datos	102
3.2.3 Dimensionamiento de la plataforma de instalación.....	102
3.2.4 Condiciones económicas y del soporte	103
3.3 Tendencias tecnológicas y de mercado	104

CAPITULO IV

4. Pruebas de verificación y control.....	106
---	-----

CAPITULO V

5. Políticas necesarias 109

CAPITULO VI

6. El éxito del Data Warehouse 114

BIBLIOGRAFIA 115

INDICE DE ILUSTRACIONES

Ilustración 2-1	6
Ilustración 2-2:	9
Ilustración 2-3	12
Ilustración 2-4	14
Ilustración 2-5	18
Ilustración 2-6	21
Ilustración 2-7	23
Ilustración 2-8	25
Ilustración 2-9	27
Ilustración 2-10	27
Ilustración 2-11	27
Ilustración 2-12	28
Ilustración 2-13	28
Ilustración 2-14	30
Ilustración 2-15	31
Ilustración 2-16	53
Ilustración 2-17	54
Ilustración 2-18	65
Ilustración 2-19	70
Ilustración 2-20	75
Ilustración 2-21	77

CAPITULO I

1. INTRODUCCIÓN

1.1 JUSTIFICACIÓN HISTÓRICA

En la actualidad, las tecnologías de la información han automatizado los procesos de carácter típicamente repetitivo o administrativo, haciendo uso de lo que se llama sistemas de información operacionales. Se entiende por aplicaciones operacionales, aquellas que resuelven las necesidades de funcionamiento de la empresa. En este tipo de sistemas, los conceptos más importantes son la actualización y el tiempo de respuesta.

Una vez satisfechas las necesidades operacionales, surge un nuevo grupo de necesidades sobre los sistemas de la empresa, a los que se califican como necesidades de información. Se entiende que son aquellas que tienen por objeto obtener la información necesaria, que sirva de base para la toma de decisiones tanto a escala estratégica como táctica. Esto se basa en gran medida en el análisis de un número extenso de datos, en el que es tan importante obtener un valor muy detallado del negocio, como el valor totalizado para el mismo. Es fundamental la visión histórica de todas las variables analizadas, y el análisis de los datos del entorno. Estos requerimientos no son, a priori, difíciles de resolver dado que la información está efectivamente en los sistemas operacionales. Cualquier actividad que realiza la empresa está reflejada de forma minuciosa en sus bases de datos.

La realidad, sin embargo, es distinta, puesto que al atender la falta de tipo informacional, los responsables de sistemas se tropiezan con múltiples problemas.

En primer lugar, al realizar consultas masivas de información (con el fin de conseguir la proporción, valor agrupado o grupo de valores solicitados), se puede ver perjudicado el nivel de servicio del resto de sistemas, dado que las consultas de las que se habla, suelen ser bastante costosas en recursos. Además, las necesidades se ven insatisfechas por la limitada flexibilidad a la hora de navegar por la información y a su inconsistencia debido a la falta de una visión global (cada visión particular del dato está almacenada en el sistema operacional que lo gestiona).

En esta situación, el siguiente paso evolutivo viene siendo la generación de un entorno gemelo del operativo, que se denomina comunmente "Centro de Información", en ésta se refresca con menor periodicidad que en los entornos operacionales y los requerimientos en el nivel de servicio al usuario son más flexibles.

Con esta estrategia se resuelve el problema de la planificación de recursos ya que las aplicaciones que precisan un nivel de servicio alto usan el entorno operacional y las que precisan consultas masivas de información trabajan en el "Centro de Información". Otro beneficio de este nuevo entorno, es la no interferencia con las aplicaciones operacionales.

Pero no terminan aquí los problemas. La información mantiene la misma estructura que en las aplicaciones operacionales por lo que este tipo de consultas debe acceder a multitud de lugares para obtener el conjunto de datos deseado. El tiempo de respuesta a las solicitudes de información es excesivamente elevado. Adicionalmente, al procesar la información de distintos sistemas, con visiones y

objetivos diferentes, en muchas ocasiones no es posible obtener la información deseada de una forma fácil y además carece de la necesaria fiabilidad.

De cara al usuario estos problemas se traducen en que no se dispone a tiempo de la información solicitada y que debe dedicarse con más intensidad a la obtención de la información que al análisis de la misma, que es donde aporta su mayor valor añadido.

1.2 SUMARIO

En los capítulos siguientes se expone, por un lado qué es un Data Warehouse, un Data Mart y el porqué de estos conceptos. Por otro lado, se estudian sus componentes de base (hardware y software) y el "*estado del arte*" de las distintas tecnologías disponibles. Se analiza las partes de las que se compone un sistema Data Warehouse y se presenta una metodología de construcción del mismo. Además se examina el uso que se le puede dar (Explotación del Data Warehouse) con especial hincapié en la Minería de Datos, y las posibilidades de acceso a esta información. También se presenta cómo algunas áreas se benefician de las tecnologías de Data Warehouse: Mercadeo, Departamento Financiero, Área de Riesgo de Crédito. Y por último, se exponen algunas políticas y recomendaciones generales a considerar para un buen uso de los sistemas de este tipo.

CAPITULO II

2. ANÁLISIS DEL ENTORNO DATA WAREHOUSE

2.1 ¿QUÉ ES UN DATA WAREHOUSE?

Tras las dificultades de los sistemas tradicionales en satisfacer las necesidades de información, surge el concepto de Data Warehouse, como solución a las necesidades de información globales de la empresa. Este término acuñado por Bill Inmon (página 3 del libro "Building the Data Warehouse"), se traduce literalmente como Almacén de Datos. No obstante si el Data Warehouse fuese exclusivamente un almacén de datos, los problemas seguirían siendo los mismos que en los Centros de Información.

La ventaja principal de este tipo de sistemas se basa en el concepto fundamental de la estructura de la información. Éste significa el almacenamiento de 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.

Un Data Warehouse es el centro de la arquitectura para los sistemas de información en los años 90. Apoya el proceso informativo proporcionando una plataforma sólida de datos integrados e históricos de la cual se hace un análisis.

Da la facilidad para la integración en un mundo de sistemas de aplicación desintegradas. Esto se logra de una manera evolutiva, un paso a la vez. Se organiza y almacena los datos necesitados para el proceso informativo y analítico sobre una

perspectiva histórica larga del tiempo. Hay de hecho un mundo de promesas en la construcción y mantenimiento de un Data Warehouse.

¿Qué es entonces un Data Warehouse?. Es una colección de datos donde siempre se cumplen las siguientes características:

- Orientados a sujetos o temas (temático)
- integrados,
- Variantes en el tiempo (histórico),
- No volátiles (permanente)

que apoyan el proceso de toma de decisiones gerenciales

Los datos que entran en el Data Warehouse vienen del ambiente operacional en casi todos los casos. Este es un almacén de datos transformados y separados físicamente de la aplicación encontrada en el ambiente operacional.

Esta definición bastante teórica merece una explicación completa porque hay algunos asuntos importantes y características que son su base. Éste se caracteriza por ser:

2.1.1 ORIENTADO A SUJETOS O TEMAS (TEMATICO)

Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Éstos 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 son más fáciles de responder dado que toda la información reside en el mismo lugar.

La primera característica de un Data Warehouse es que está orientado alrededor de los temas principales de la empresa. En el manejo de los datos, la orientación de temas, esta en contraposición a la orientación clásica de aplicaciones proceso/funcional, lo cuales, la mayoría de los mas antiguos sistemas operacionales están organizados de esa forma.

La ilustración 2-1 muestra el contraste entre los dos tipos de orientaciones.

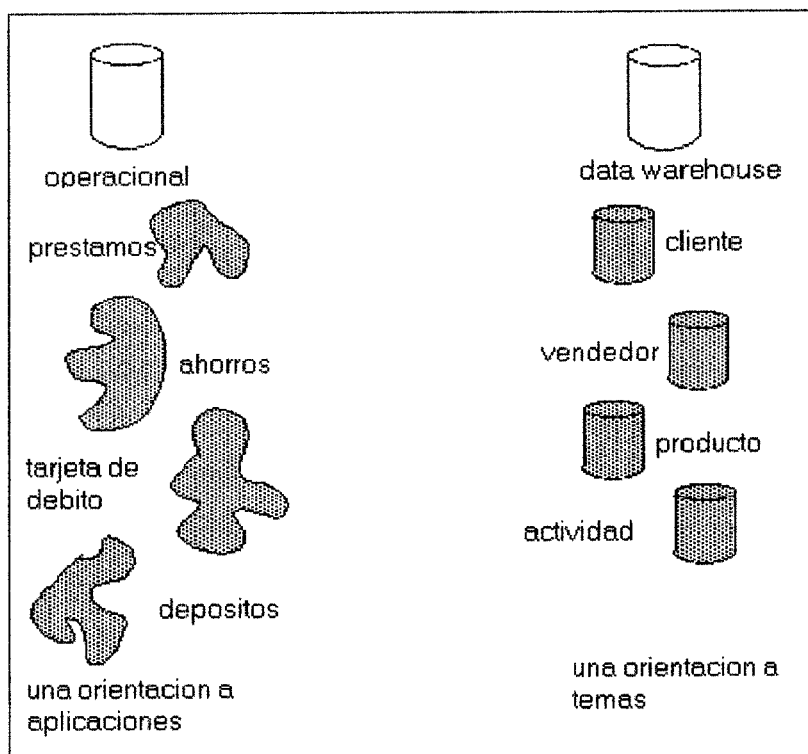


Ilustración 2-1

El mundo operacional se diseña alrededor de aplicaciones y de funciones tales como préstamos, ahorros, tarjeta de débito y crédito para una institución

financiera. El Data Warehouse se ordena alrededor de temas importantes tales como cliente, vendedor, producto y actividad. La alineación de estos afecta el diseño y la implantación de los datos encontrados en el Data Warehouse. Lo que destaca, es que la mayoría de áreas sobre temas influyen la parte importante de la estructura clave.

El mundo de la aplicación se refiere al diseño de base de datos y al diseño de procesos. El Data Warehouse se enfoca en el modelado de los datos y el diseño de la base de datos exclusivamente. El diseño de procesos (en su forma clásica) no es parte del ambiente de éste.

Las diferencias entre una aplicación orientada a procesos y una orientada hacia temas muestran una diferencia adicional en el contenido de los datos a un nivel detallado.

El Data Warehouse excluye los datos que no son utilizados para el procesamiento en la toma de decisiones, mientras que las aplicaciones orientadas a operaciones contienen datos para satisfacer los requisitos inmediatos de funcionalidad y procesamiento que pueden o no ser de usados en la toma de decisiones.

Otra diferencia importante está en las relaciones de los datos. El ambiente operacional mantiene una relación en curso entre dos o más tablas basadas en una regla de negocio que esté en efecto. El Data Warehouse abarcan un espectro del tiempo y las relaciones encontradas en éste son muchas. Muchas reglas de negocio (y correspondientemente, muchas relaciones de los datos) se representan en éste entre dos o más tablas.

De ninguna otra perspectiva que el de la diferencia fundamental entre una aplicación con orientación funcionalidad/proceso y una orientación a temas o sujetos, hay una diferencia importante entre los sistemas y los datos operacionales y el Data Warehouse.

2.1.2 INTEGRADO

Sencillamente, el aspecto más importante del ambiente del Data Warehouse es que los datos encontrados dentro del Data Warehouse están integrados. **Siempre y sin excepciones.** La misma esencia del ambiente del Data Warehouse es que los datos contenidos dentro de los límites del almacén están integrados.

La integración se muestra en muchas diversas maneras - en convenciones de nombramiento constantes, en la medida constante de variables, en estructuras de codificación constantes, en atributos físicos constantes de datos, y así sucesivamente.

Ponga en contraste la integración encontrada dentro del Data Warehouse con la carencia de la integración en el ambiente de las aplicaciones, y las diferencias son rígidas, según lo mostrado por la ilustración 2-2 (pagina 9) en la que se muestran algunas de las mas importantes diferencias en que las aplicaciones son diseñadas.

A lo largo de los años los diversos diseñadores de aplicaciones toman decisiones individuales numerosas en cuanto a cómo una aplicación debe ser construida. El estilo y las decisiones individualizadas del diseño por parte del

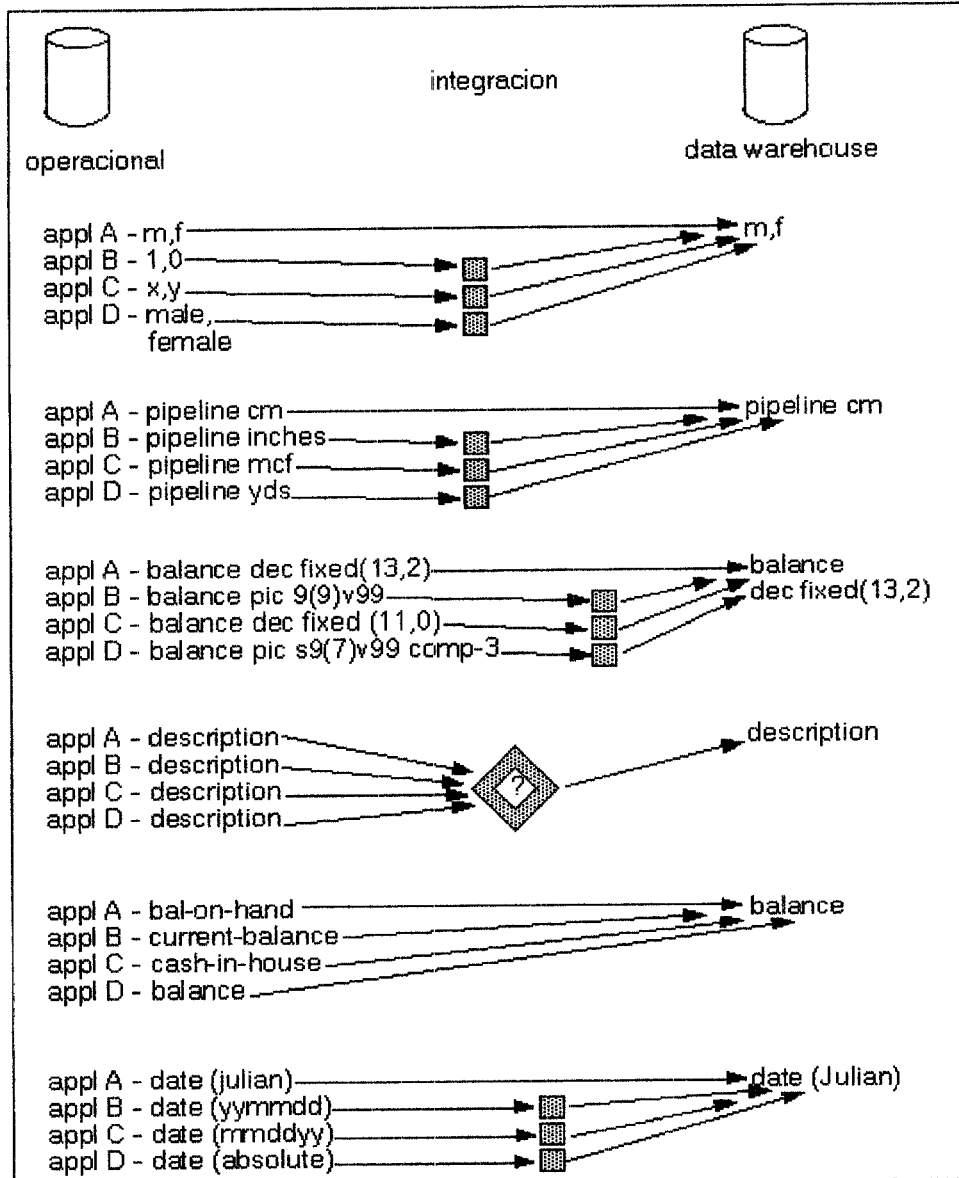


Ilustración 2-2.

diseñador de la aplicación se muestran en muchas maneras, la codificación, las estructuras clave, características físicas, en las convenciones de nombramiento y así sucesivamente. La capacidad colectiva de muchos diseñadores en crear aplicaciones inconsistentes es bien conocida.

2.1.2.1 CODIFICACIÓN

Se elige codificar el campo GÉNERO de diversas maneras. Un diseñador representa el sexo de una persona como un " M " y " F. " Otro lo hace con un " 1 " y un " 0. " Otro lo representa como un " x " y " y. " Pero otro lo define como " masculino " y " femenino. " No importa mucho cómo llegue en el Data Warehouse. " M " y " F " son probablemente tan buenos como cualquier representación. Qué importa cualquier fuente de donde viene, debe llegar a éste en un estado integrado constante. Por lo tanto cuando se carga desde una aplicación donde se ha representado de otra forma que un formato de " M " y de " F ", los datos se deben convertir al formato del Data Warehouse.

2.1.2.2 MEDIDA DE LOS ATRIBUTOS

Los diseñadores de aplicaciones miden las distancias en una variedad de maneras a través de los años. Éste almacena datos de las distancias en centímetros. Otro los guarda en términos de pulgadas. Otro los almacena en metros. Y otro en términos de yardas. Cualquiera que sea la fuente, cuando la información de la distancia llega al Data Warehouse se necesita ser medida de la misma manera.

Según lo mostrado en la ilustración 2-2 (página 9), las tareas de integración afectan casi cada aspecto del diseño - las características físicas de los datos, el dilema del tener más de una fuente de datos, la edición de estándares de nombramiento inconsistentes, formatos de fecha contrarios, y así sucesivamente.

Cualquiera que sea el diseño, el resultado es igual - los datos necesitan ser salvados en el Data Warehouse en una manera singular, de aceptación global e incluso cuando los sistemas operacionales subyacentes almacenan los datos diferentemente.

Cuando el analista de sistemas mira el Data Warehouse, la concentración debe estar en usar los datos que están en el almacén, que en preguntarse sobre la credibilidad o el estado coherente de los datos.

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.

2.1.3 VARIANTE EN EL TIEMPO (HISTORICO)

Todos los datos en el Data Warehouse son exactos en determinado momento en el tiempo. Esta característica básica de los datos en el almacén es muy diferente de los datos encontrados en el ambiente operacional. En éste son

exactos en el momento del acceso. Es decir, cuando se tiene acceso a una unidad de datos, se cuenta con que refleje valores exactos en ese momento.

Porque los datos en el Data Warehouse son exactos en un cierto momento en el tiempo (es decir, "no ahora"), los datos encontrados en el almacén se dicen venir a ser la "variante del tiempo."

La ilustración 2-3 muestra la variación del tiempo de los datos del Data Warehouse.

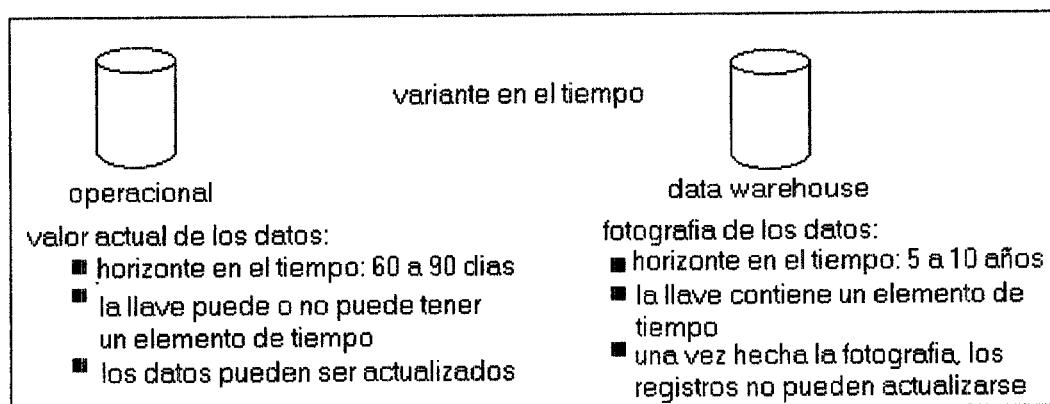


Ilustración 2-3

La variación del tiempo de los datos del Data Warehouse se muestra de varias maneras. La más simple es que se representan sobre un horizonte largo del tiempo - desde cinco hasta diez años. El horizonte del tiempo representado para el ambiente operacional es mucho más corto - de los valores actuales de hoy hasta sesenta a noventa días. Las aplicaciones que deben realizarse bien y deben estar disponibles para el tratamiento transaccional deben llevar la cantidad mínima de datos si tienen cualquier grado de flexibilidad del todo. Por lo tanto las

aplicaciones operacionales tienen un horizonte corto del tiempo, como cuestión de diseño sano de la aplicación.

La segunda manera que se muestra es en la estructura clave. Cada una de éstas contiene - implícita o explícitamente - un elemento del tiempo, tal como día, semana, mes, año. Éste es casi siempre la clave de fondo encontrado en el Data Warehouse. Ocasionalmente, existe implícito, por ejemplo el caso donde un archivo entero se duplica en el final del mes, o en el trimestre.

La tercera manera que aparece la variante en el tiempo es en los datos del Data Warehouse, una vez que estén grabados correctamente, no pueden ser actualizados. Éstos son, para todos los propósitos prácticos, una serie larga de "fotos de los datos". Por supuesto si éstas se han tomado incorrectamente, entonces las fotos pueden ser cambiadas. Pero si se asume que las fotos de los datos están correctamente, estas no se alteran una vez hechas. En algunos casos puede ser poco ético o ilegal que las fotos de los datos sean alteradas. Los datos operacionales, siguen siendo exactos en el momento del acceso; pueden ser actualizados al surgir la necesidad.

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, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

2.1.4 NO VOLATIL

La cuarta característica en definir es que no es volátil (permanente). La ilustración 2-4 enseña este aspecto.

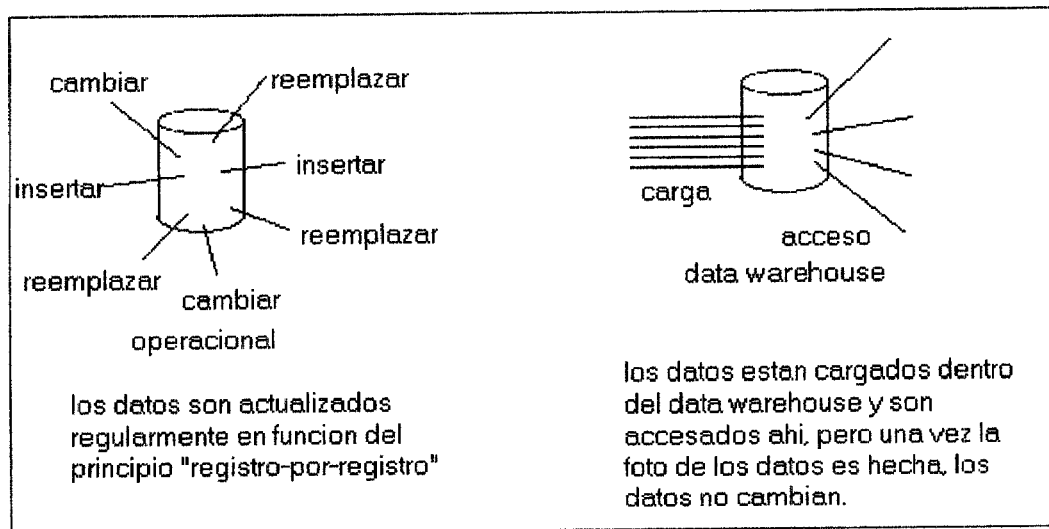


Ilustración 2-4

En ésta se muestra que las actualizaciones - inserciones, eliminaciones, y cambios - están hechas regularmente al ambiente operacional en un principio de "registro-por-registro". Pero la manipulación básica de los datos que sucede en el Data Warehouse es mucho más simple. Hay solamente dos clases de operaciones que ocurren - la carga inicial de datos, y el acceso de los datos. No hay actualización (en el sentido general de la actualización) como parte normal de proceso.

Hay algunas poderosas consecuencias de esta diferencia básica entre el procesamiento operacional y el procesamiento en el Data Warehouse. En el nivel del diseño, la necesidad de ser cauteloso por anomalías en la actualización no es ningún factor en éste, puesto que la actualización de datos no se hace. Esto

significa eso en el nivel físico del diseño, se puede tomar libertades para optimizar el acceso de datos, particularmente en atender las cuestiones de normalización y desnormalización. Otra consecuencia de la simplicidad de la operación del Data Warehouse está en la tecnología subyacente usada para ejecutar el ambiente de éste. Teniendo que utilizar la actualización registro-por-registro en un modo en línea (al igual que el caso con el proceso operacional) se requiere la tecnología para tener una fundación muy compleja por debajo de una fachada de la simplicidad. Esto debe soportar respaldo y recuperación, la integridad de las transacciones y de los datos, y la detección y solución de bloqueos son absolutamente complejos e innecesarios para el procesamiento en el Data Warehouse.

Las características - diseño de orientación a temas, integración de datos, la variación en el tiempo, y la simplicidad de la administración de datos - todas conducen a un ambiente que es **muy** diferente del ambiente operacional clásico.

La fuente de casi todos los datos del Data Warehouse es el ambiente operacional. Es una tentación pensar que hay redundancia masiva de datos entre los dos ambientes. De hecho, la primera impresión que mucha gente deriva es la de la gran redundancia de datos entre el ambiente operacional y el ambiente de Data Warehouse. Tal comprensión es superficial y demuestra una carencia de entender en cuanto a que está ocurriendo en el Data Warehouse. De hecho hay un **mínimo** de redundancia de datos entre el ambiente operacional y los datos del Data Warehouse.

Se considera lo siguiente:

- Se filtran los datos mientras que pasan del ambiente operacional al ambiente del Data Warehouse. Muchos datos nunca pasan de éste. Solamente esos datos que son necesarios encuentran su camino en el ambiente del Data Warehouse.
 - El horizonte del tiempo de datos es muy diferente a partir de un ambiente al siguiente. Los datos en el ambiente operacional están muy frescos. Los datos en el almacén son mucho más antiguos. Desde la perspectiva de los horizontes en el tiempo solamente, hay muy poco traslape entre los ambiente operacional y Data Warehouse.
 - El Data Warehouse contiene los datos resumidos que nunca se encuentran en el ambiente operacional.
 - Los datos experimentan una transformación fundamental mientras que pasan dentro del Data Warehouse. La ilustración 2-1 (página 6) muestra que la mayoría de los datos están alterados perceptiblemente por encima de ser seleccionados y movidos dentro del Data Warehouse. Dicho otra manera, la mayoría de los datos se altera físicamente y radicalmente mientras se mueve dentro del almacén. No son los mismos datos que residen en el ambiente operacional a partir del punto de vista de la integración.
-

A la luz de estos factores, la redundancia de datos entre los dos ambientes es una ocurrencia rara, dando por resultado menos del 1% de redundancia entre los dos ambientes.

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

2.2 LA ESTRUCTURA DEL DATA WAREHOUSE

2.2.1 ALMACENAMIENTO

Un Data Warehouse tiene una estructura distinta. Hay diversos niveles de resumen y detalle que lo demarcan. La estructura de un Data Warehouse es mostrada por la ilustración 2-5 (página 18).

Se observan los diversos componentes del Data Warehouse:

- Meta datos,
- Datos detallados actuales,
- Datos detallados más antiguos,
- Datos ligeramente resumidos, y
- Datos altamente resumidos.

En gran medida la preocupación principal son los “datos detallados actuales”. Esto se debe porque: los “datos detallados actuales” reflejan los sucesos más recientes, que son siempre de gran interés; son voluminosos

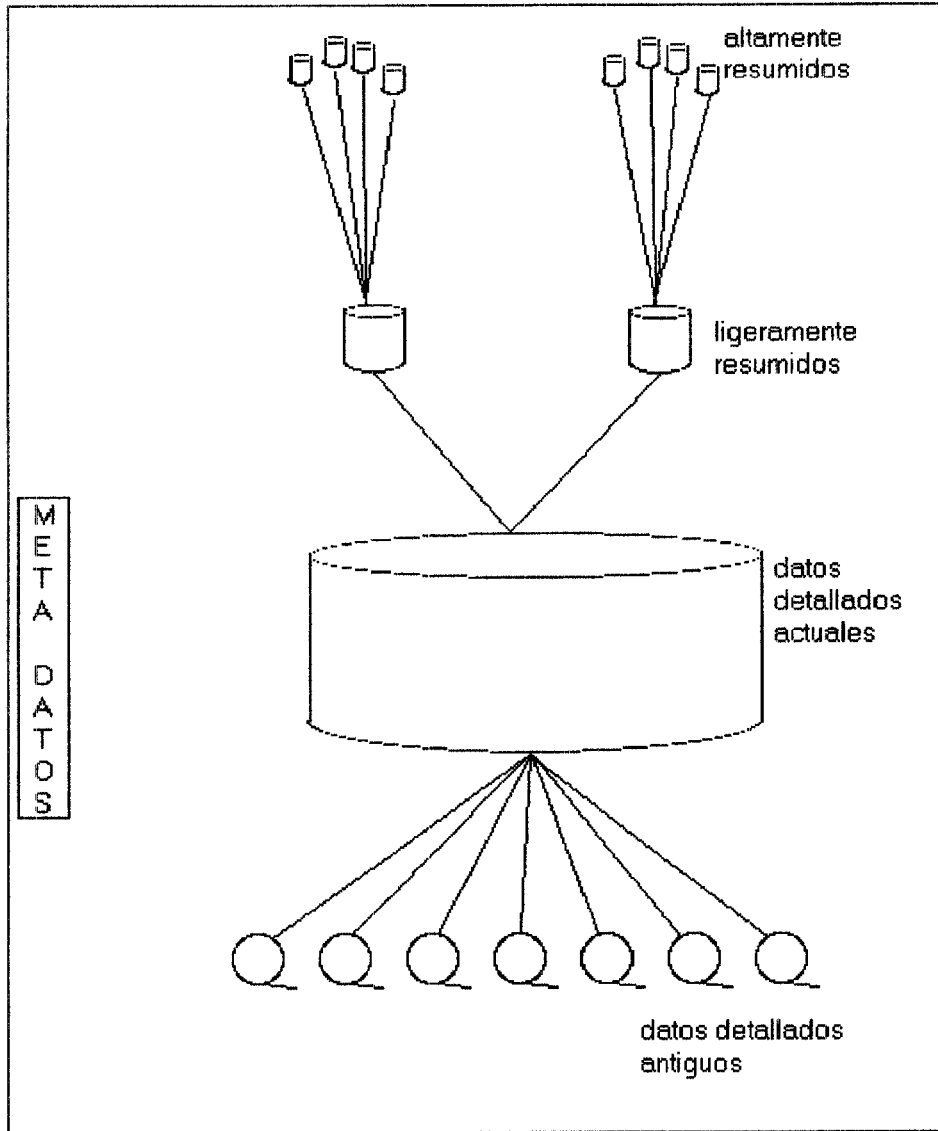


Ilustración 2-5

porque se almacenan en el nivel más bajo de granularidad ; y se guardan casi siempre en discos, al cual es rápido tener acceso, pero costoso y complejo para manejar.

Los “datos detallados más antiguos” son los que se guardan en una cierta forma de almacenamiento en masa. No son frecuentemente accesados y almacenados en un nivel desglosado consistente con “datos detallados actuales”. Mientras que no es obligatorio que se guarden en un medio de almacenamiento alterno, debido al gran esperado volumen de datos unido con el acceso infrecuente de éstos, el medio de almacenamiento para los “datos detallados más antiguos” no es generalmente en discos.

Los “datos ligeramente resumidos” son los que se destilan desde el nivel bajo del detalle encontrado a el nivel detallado actual. En este se guarda casi siempre en discos. Las ediciones del diseño para el arquitecto de los datos en la construcción de este nivel del Data Warehouse son:

- Qué unidad del tiempo vuelve a hacer el resumen, y
- Qué contenido - atributos - incluyen los “datos ligeramente resumidos”.

El nivel siguiente de datos encontrados en el Data Warehouse es el de los “datos altamente resumidos”. Éstos son compactos y fácilmente accesibles. Además se encuentran a veces en el ambiente del Data Warehouse y en otros casos se encuentran fuera de las barreras inmediatas de la tecnología que contiene el Data Warehouse. (En cualquier caso, éstos son parte del Data Warehouse sin importar donde los datos se contienen físicamente.)

El componente final del Data Warehouse es el de meta datos. De muchas maneras éstos se colocan en una dimensión diferente que otros datos del Data Warehouse, porque los meta datos no contienen ningún dato tomado directamente del ambiente operacional. Desempeñan un papel especial y muy importante porque se utilizan como:

- Un directorio para ayudar al analista del Sistemas a localizar el contenido del Data Warehouse,
- Una guía para mapeo de datos de como los datos se transforman del ambiente operacional al ambiente del Data Warehouse,
- Una guía de los algoritmos usados para el resumen entre los “datos detallados actuales” y los “datos ligeramente resumidos”, y entre los “datos ligeramente resumidos” y los “datos altamente resumidos”.

Los meta datos desempeñan un papel mucho más importante del que jamás se hizo en el sistema operacional clásico.

Para traer a la vida los diversos niveles de datos encontrados en el Data Warehouse, considere el ejemplo mostrado en la ilustración 2-6 (pagina. 21).

Este detalle antiguo sobre depósitos es anterior al año 1992. Todo el detalle de los depósitos a partir de 1982 (o siempre que el arquitecto de los datos pueda comenzar a recoger el detalle perteneciente al archivo) se guarda en el nivel de “datos detallados más antiguos”.

El detalle actual de valores contiene datos a partir de 1992 a 1993 (si se asume que 1993 es el año actual). En general los depósitos no encuentran su

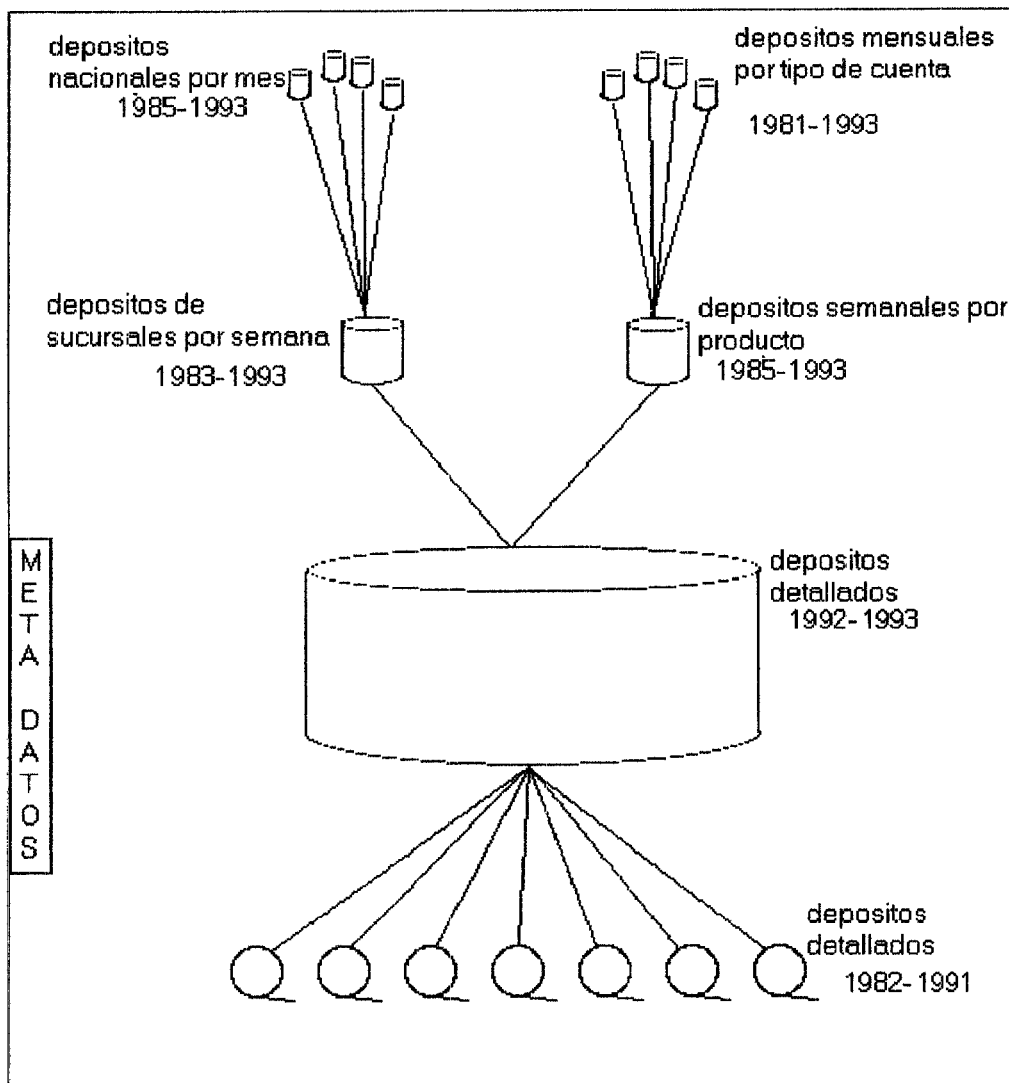


Ilustración 2-6

forma en el nivel actual hasta que por lo menos pasen veinticuatro horas desde que la información de los depósitos llegó a estar disponible para el ambiente operacional. Es decir que hay un retraso de tiempo por lo menos de veinticuatro horas entre que el ambiente operacional consigue las noticias de los depósitos y del momento en que estos datos fueron incorporados en el Data Warehouse.

El detalle de los depósitos es resumido semanalmente por la línea de productos y por las sucursales o agencias para producir los almacenes ligeramente resumidos de datos.

El detalle semanal de los depósitos es más a fondo resumido mensualmente incluso a lo largo de líneas más generales para producir los datos altamente resumidos.

Los meta datos contienen por lo menos:

- la estructura de los datos,
- los algoritmos usados para el resumen,
- la asociación del ambiente operacional al Data Warehouse.

Cada resumen que se hace, no siempre consigue grabarse en el Data Warehouse. Hay ocasiones donde el análisis es hecho y un tipo o el otro de resumen es producido. El único tipo que se guarda permanentemente es el de los datos que se utilizan con frecuencia. Es decir si un analista del sistemas produce un resultado resumido que tenga una probabilidad muy baja de ser utilizado otra vez, entonces ese resumen no se guarda en el Data Warehouse.

2.2.2 FLUJO DE DATOS

Hay un flujo normal y predecible de datos dentro del Data Warehouse. La ilustración 2-7 lo muestra.

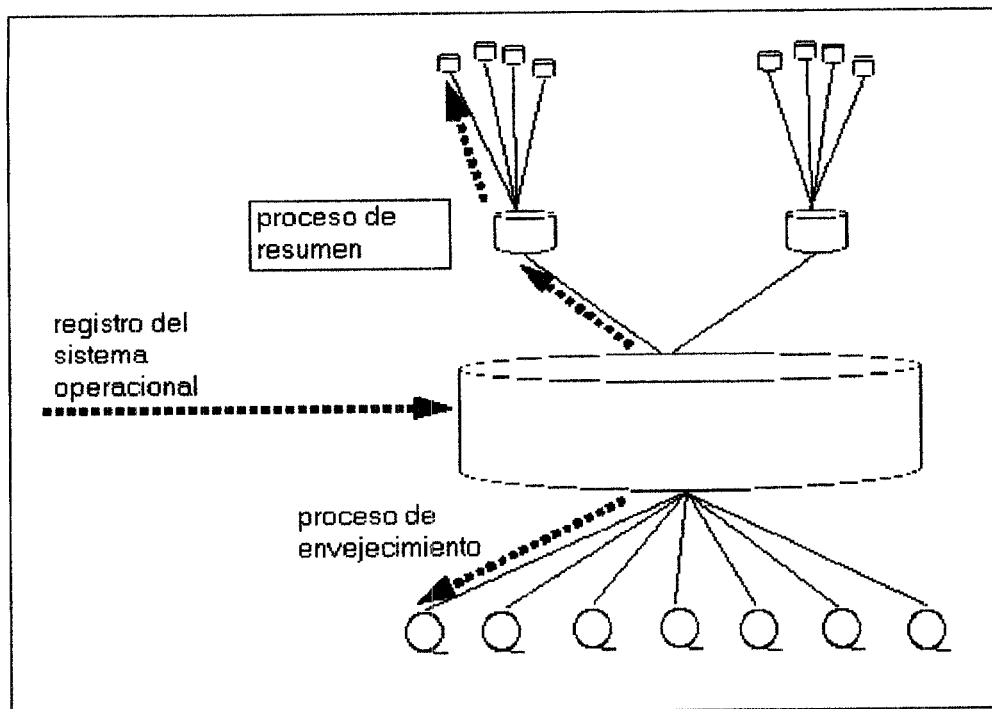


Ilustración 2-7

Los datos entran desde el ambiente operacional.

NOTA: Hay algunas anomalías muy interesantes a esta regla. Sin embargo, **casi todos los datos** entran desde el ambiente operacional.

Mientras los datos entran, se transforman, como se describe anteriormente. Al entrar en el nivel de detalle actual, según lo mostrado, residen allí y se utilizan hasta que ocurre uno de tres acontecimientos:

- Se depura,
- Se resume, y/o
- Se archiva.

El proceso del envejecimiento dentro de un Data Warehouse mueve los “datos detallados actuales” a “datos detallados mas antiguos”, basados en la edad de datos. El proceso de resumen utiliza el detalle de datos para calcular los datos ligeramente resumidos y los datos altamente resumidos.

Hay algunas excepciones al flujo según lo mostrado. Sin embargo, en general, para la mayoría extensa de datos encontrados dentro de un Data Warehouse, el flujo de datos está según lo representado.

2.2.3 UTILIZACIÓN DEL DATA WAREHOUSE

Los diversos niveles de datos dentro del Data Warehouse reciben diversos usos. En general, cuanto más alto es el nivel del resumen, más son los datos utilizados, mientras que los datos detallados más antiguos casi apenas se utilizan según lo mostrado en la ilustración 2-8 (pagina 25).

Hay una buena razón para mover la organización al paradigma sugerido en ésta: utilización de recursos. Cuanto más resumidos los datos, más rápido y más eficiente es conseguirlos. Si un departamento encuentra que está haciendo mucho procesamiento en los niveles detallados del Data Warehouse, entonces una cantidad grande de recursos correspondientemente de la máquina se está

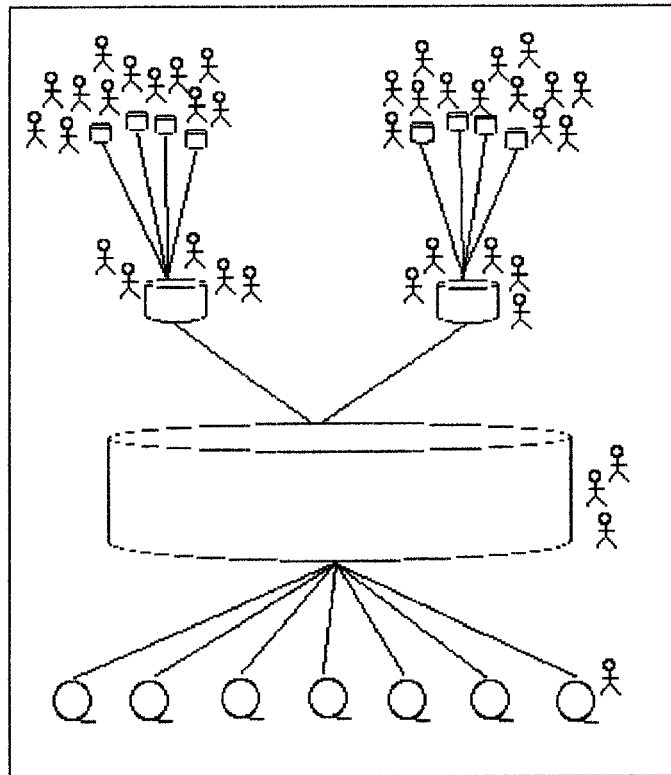


Ilustración 2-8

consumiendo. Esta en cada uno hacer el mejor interés de procesar en un alto nivel del resumen como sea posible.

Para muchos departamentos, el analista de sistemas en un ambiente pre-Data Warehouse utiliza datos en el nivel detallado. El conseguir éstos es como una manta de seguridad, incluso cuando otros niveles del resumen están disponibles. Hay dos motivadores en la disposición del arquitecto de los datos:

- Instalando un sistema de "cargos de vuelta", donde el usuario final paga los recursos consumidos, y
- Precisando que un tiempo de respuesta muy bueno puede ser alcanzado cuando se referencia a datos en un alto nivel de resumen, mientras que el tiempo de respuesta pobre resulta de ocuparse de datos en un nivel bajo de detalle.

E.F. Codd, considerado como el padre de las bases de datos relacionales, viene insistiendo desde principio de los noventa, que disponer de un sistema de bases de datos relacionales, no significa disponer de un soporte directo para la toma de decisiones (página 6 del libro "Oracle Data Warehouseing"). Muchas de estas decisiones se basan en un análisis de naturaleza multidimensional, que se intenta resolver con la tecnología no orientada para esta naturaleza. Este análisis multidimensional, parte de una visión de la información como dimensiones de negocio.

Éstas se comprenden mejor fijando un ejemplo, para un sistema de caja, las jerarquías que se podrían manejar para las dimensiones: zona geográfica, tipo de producto y tiempo en que se hizo la transacción. La visión general de la

información de depósitos para estas dimensiones definidas, se representan en la ilustración 2-9

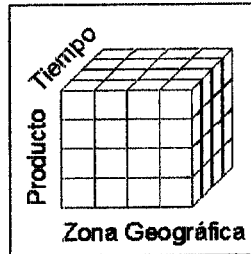


Ilustración 2-9

Un gerente de una agencia bancaria puede estar interesado en visualizar la información para su ubicación en el tiempo para todos los productos que ofrece, lo podría tener una representación gráfica como la ilustración 2-10

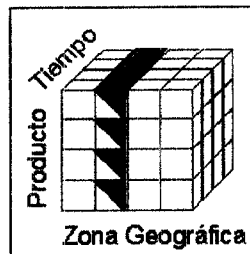


Ilustración 2-10

Un gerente de producto, sin embargo desea examinar la distribución geográfica de sus productos, para toda la información histórica almacenada en el Data Warehouse. Esto se podría representar como la ilustración 2-11

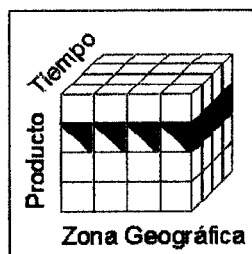


Ilustración 2-11

O se podría también estudiar los datos en un determinado momento o una visión particularizada. Esto se muestra en la ilustración 2-12



Ilustración 2-12

A su vez estas dimensiones tienen una jerarquía, interpretándose en el cubo como que cada cubo elemental es un dato elemental, del que se puede extraer información agregada. Lo anterior se muestra en la ilustración 2-13:

Zona Geográfica	Ciudad	Agencia 1 Agencia 2 Agencia 3	
	Occidente	Agencia Departamental 1 Agencia Departamental 2	
Producto	Ahorros	Con sorteo Con intereses	Intereses Diarios Intereses Mensuales
	Monetarios	Corriente Con intereses	
	Plazo Fijo	A 30 días A 60 días A 90 días	
Tiempo	Año 1999	1er trimestre	Ene/1999 Feb/1999 Mar/1999
	Año 2000	2do Trimestre	

Ilustración 2-13

Y así se puede querer analizar la evolución de los depósitos en la agencia 1 para las cuentas de ahorro con intereses diarios en los meses de Enero a Marzo de 1999.

Ello es fácil de obtener si la información de depósitos se ha almacenado en un Data Warehouse, definiendo estas jerarquías y estas dimensiones del negocio.

Otra característica del Data Warehouse es que contiene datos relativos a los datos, concepto que se ha venido asociando al término de meta datos. Éstos permiten mantener registro de la procedencia de la información, la periodicidad de refresco, su fiabilidad, forma de cálculo relativa a los datos de nuestro almacén.

Estos meta datos son los que permitan simplificar y automatizar la obtención de la información desde los sistemas operacionales a los sistemas de información.

Los objetivos que deben cumplir los meta datos, según el grupo al que va dirigido, son:

- Soporte al usuario final, ayudándole a acceder al Data Warehouse con su propio lenguaje de negocio, indicando qué información hay y qué significado tiene. Ayudar a construir consultas, informes y análisis, mediante herramientas de navegación.
 - Soporte a los responsables técnicos del Data Warehouse en aspectos de auditoría, gestión de la información histórica, administración del Data Warehouse, elaboración de programas de
-

extracción de la información, especificación de las interfaces para la realimentación a los sistemas operacionales de los resultados obtenidos.

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

La ilustración 2-14 muestra gráficamente estos procesos.

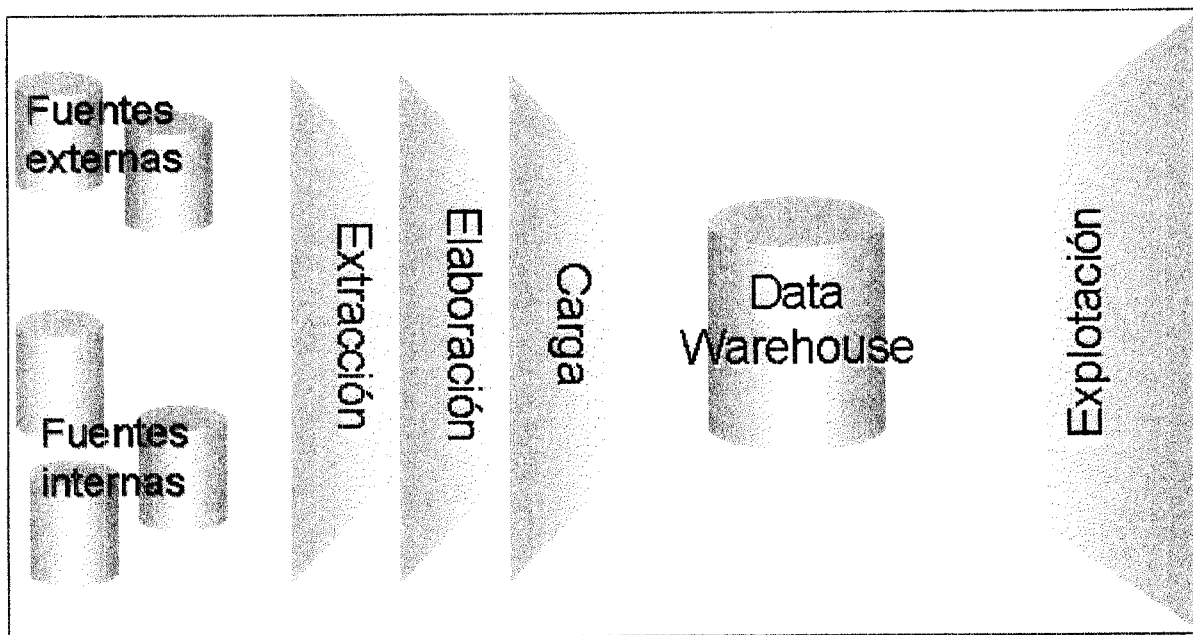


Ilustración 2-14

A continuación se describen dichos procesos clave en la gestión de un Data Warehouse:

- Extracción: obtención de información de las distintas fuentes tanto internas como externas.
- Elaboración: filtrado, limpieza, depuración, homogeneización y agrupación de la información.

- Carga: organización y actualización de los datos y los meta datos en la base de datos.
- Explotación: extracción y análisis de la información en los distintos niveles de agrupación.

Desde el punto de vista del usuario, el único proceso visible es la explotación del almacén de datos, aunque el éxito del Data Warehouse radica en los tres procesos iniciales que alimentan la información del mismo y suponen el mayor porcentaje de esfuerzo (en torno a un 80%) a la hora de desarrollar el almacén.

Las diferencias de un Data Warehouse con un sistema tradicional se pueden resumir en la ilustración 2-15:

SISTEMA TRADICIONAL	DATA WAREHOUSE
• Predomina la actualización	• Predomina la consulta
• La actividad más importante es de tipo operativo (día a día)	• La actividad más importante es el análisis y la decisión estratégica
• Predomina el proceso puntual	• Predomina el proceso masivo
• Mayor importancia a la estabilidad	• Mayor importancia al dinamismo
• Datos en general desagregados	• Datos en distintos niveles de detalle y agregación
• Importancia del dato actual	• Importancia del dato histórico
• Importante del tiempo de respuesta de la transacción instantánea	• Importancia de la respuesta masiva
• Estructura relacional	• Visión multidimensional
• Usuarios de perfiles medios o bajos	• Usuarios de perfiles altos
• Explotación de la información relacionada con la operativa de cada aplicación	• Explotación de toda la información interna y externa relacionada con el negocio

Ilustración 2-15

Una de las claves del éxito en la construcción de un Data Warehouse es el desarrollo de forma gradual, seleccionando a un departamento usuario como piloto y expandiendo progresivamente el almacén de datos a los demás usuarios. Por ello es importante elegir este usuario inicial o piloto, siendo importante que sea un departamento con pocos usuarios, en el que la necesidad de este tipo de sistemas es muy alta y se puedan obtener y medir resultados a corto plazo.

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 modelado 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 relacionados 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.
-

2.3 CONCEPTOS Y FUNCIONALIDADES BÁSICAS

2.3.1 LA FUNCIÓN BANCARIA

Para establecer cual es la función de la banca en nuestro país, nos referimos a un extracto de la Ley de Bancos y Grupos Financieros (Decreto Número 315 del Congreso de la República de Guatemala)

Titulo IV

LOS BANCOS Y SUS OPERACIONES Y SERVICIOS

Capítulo Único

Artículo 43. Bancos Universales. Los bancos universales son entidades financieras autorizadas conforme esta ley, que tienen por objeto realizar todas las operaciones y prestar todos los servicios financieros permitidos a las entidades que conforman los grupos financieros, excepto las operaciones de banco de inversión.

Artículo 44. Bancos Comerciales. Los bancos comerciales son entidades financieras autorizadas conforme esta ley, que tienen por objeto captar recursos del público y canalizarlos al financiamiento de las operaciones indicados en la presente ley y prestar servicios financieros que sean compatibles con su naturaleza.

Los bancos comerciales podrán realizar las operaciones y servicios siguientes:

- a) Captar recursos del publico mediante depósitos de cualquier clase en moneda nacional y extranjera, emisión y colocación de bonos, empréstitos en el país o en el extranjero, emisión y colocación de
-

obligaciones convertibles en acciones. En el caso de las captaciones en moneda extranjera deberán realizarlas de conformidad con las disposiciones que para el efecto emita la junta monetaria;

- b) Otorgar crédito a corto, mediano y largo plazo, con garantías personales, hipotecarias, prendarias o combinadas;
- c) Invertir en valores emitidos y/o garantizados por el Estado o sus instituciones, por los bancos autorizados conforme esta ley o por empresas privadas, siempre que los valores de estas últimas, estén calificados, como mínimo en categoría A, por una empresa internacional calificadora de riesgo de reconocido prestigio. La Junta Monetaria podrá autorizar la inversión en valores en moneda extranjera; y,
- d) Prestar servicios financieros y efectuar operaciones de confianza acordes a la naturaleza de sus operaciones;

Artículo 45. Bancos de Inversión. Los Bancos de Inversión son entidades financieras autorizadas conforme esta ley, que tienen por objeto promover la inversión en general para el desarrollo de la producción nacional, actuando como inversionistas directos e indirectos en empresas productivas e intermediarios entre inversionistas y emisores de deuda.

Los bancos de inversión financiarán sus actividades con su propio capital y con recursos provenientes de la emisión y colocación de títulos en oferta pública bursátil, siempre que el plazo de los títulos negociados no sea menor de un año, que los mismos no sean susceptibles de redención anticipada, que sean

emitidos en serie y que los títulos tengan las mismas características cuando formen parte de la misma serie.

Los Bancos de Inversión podrán realizar las operaciones activas y de confianza siguientes:

- a) Canalizar recursos a plano no menor de tres (3) años mediante el otorgamiento de financiamiento para la creación, desarrollo, transformación y fusión de empresas productivas;
- b) Suscribir, adquirir, mantener en cartera y negociar valores y acciones de las empresas a que se refiere el inciso anterior, siempre que se trate de sociedades anónimas;
- c) Otorgar financiamiento a plano menor de tres (3) años a las empresas en las que tenga una participación mínima del veinticinco por ciento (25%) del capital pagado; o, en aquellas a las cuales les haya otorgado financiamiento a plazo mayor de tres (3) años, hasta por el veinticinco por ciento (25%) de dicho financiamiento, siempre que no exceda los límites establecidos en la presente ley; y,
- d) Prestar servicios de agente financiero en Bolsa de Valores.

2.3.2 ASPECTOS A CONSIDERAR

2.3.2.1 CAMBIO DE HÁBITOS EN LA TOMA DE DECISIONES

Como primer medida, la aproximación para la toma de decisiones puede provocar un cambio diario en las compañías dedicadas a los

negocios. Muchas organizaciones están a mitad de camino entre una expansión y una disminución; muchas están experimentando con soluciones cliente/servidor que potencian su equipo para hacer decisiones documentadas. Los usuarios realizan sus negocios diarios usando un conjunto predeterminado de datos impresos en papel o en la pantalla, el cual les ayuda a realizar una gestión actual, histórica y futura. Estos sistemas permiten que dicha gestión analice la salida de los informes del sistema y haga decisiones inteligentes para el futuro del negocio. Muchos sistemas heredados carecen de una característica importante que restringe la capacidad para hacer este tipo de decisiones; el usuario no posee un mecanismo robusto que le proporcione datos ajustados y exactos para facilitar este proceso. Las siguientes tres áreas clave resumen las imperfecciones de muchos sistemas de referencia actuales:

- Accesibilidad a los datos – con los sistemas tradicionales, el usuario debe utilizar los servicios técnicos del departamento de informática para acceder a los datos deseados--. Los usuarios no pueden conseguir lo que quieren cuando lo desean.
 - Puntualidad en la ejecución de la petición –los servicios de información centralizados deben proporcionar soluciones siguiendo un proceso primero en llegar, primero en servirse (FCFS), y son incapaces de manejar peticiones inmediatas de forma adecuada.
-

- Formato --las impresiones en papel producidas por la mayoría de los sistemas de referencia tradicionales son inútiles para los usuarios que necesitan un diagrama, una hoja de cálculo o una presentación visual de más alto nivel.

Cuando se inicia un sistema Data Warehouse, las corporaciones necesitan una clara visión de lo que están planificando para completar la iniciativa. Las decisiones de diseño se deben hacer pensando en los objetivos, así como en una visión sobre cómo ayuda éste a la compañía. Las áreas del negocio aisladas y específicas de la organización que suelen tener los usuarios finales forman parte del nuevo sistema. Estas deben satisfacer oportunamente las peticiones de datos, que son necesarios para hacer importantes. Las dos áreas que plantean un mayor riesgo son:

- Resistencia a los cambios --el equipo debe incluir personal de la compañía encomendado a la iniciativa y debe observar el valor real de un cambio hacia una solución de información para toma de decisiones.
- Fallos del equipo a la hora de escuchar y buscar la participación de los clientes del Data Warehouse.

No se debe cambiar por el simple hecho de cambiar. Esto nunca funciona, y tampoco en un Data Warehouse. Una parte del éxito en el desarrollo de éste consiste en obtener y dar los resultados (aunque parezcan pequeños) de forma oportuna a la comunidad de usuarios. Se debe animar al usuario a tomar decisiones de alto nivel en el negocio basándose en la

información recogida de un Data Warehouse, aunque pierda el interés si no dispone de una oportunidad para hacerlo.

2.3.2.2 CAMBIO EN LA FORMA DE DESARROLLAR SOLUCIONES DE LA TECNOLOGÍA DE LA INFORMACIÓN

Los proyectos de Data Warehouse requieren una mayor comunicación. Para los nuevos programadores, esto debe ser fácil. Para un veterano, esto puede ser difícil o imposible. Éstos no son para todos. ¿Qué es lo que tanto diferencia al desarrollo con respecto a la metodología de los clásicos sistemas operacionales?

- Los módulos se pasan a la comunidad de usuarios por trozos. En el modo de desarrollo de los clásicos sistemas operacionales, éstos se unen y se distribuyen como una sola unidad después de pasar por las fases de garantía de calidad y de aceptación del usuario. En el Data Warehouse, los usuarios pueden recibir los primeros resultados según se introducen los datos en el almacén.
 - Cuando se desarrolla un sistema operacional, el equipo del proyecto posee una copia de los requisitos y las funcionalidades exigidas por los usuarios y no se puede desviar del plan previamente acordado. No se sugiere que el equipo se olvide de la planificación de proyectos o de la documentación, pero sus miembros deben ser más flexibles.
-

Una gran diferencia entre un Data Warehouse y los proyectos de los sistemas operacionales es el interfaz con la comunidad de usuarios. Éstos documentos, normalmente se realizan antes de escribir una sola línea de código, definen la funcionalidad y la naturaleza de los módulos del sistema. Se debe pactar previamente si el desarrollo es propio o es externo, reorientando su naturaleza o incorporando nuevas funcionalidades cuando el sistema se están desarrollando.

2.3.2.3 ALTA INTEGRACIÓN ENTRE LOS COMPONENTES

En los clásicos sistemas operacionales, a veces, existe un diálogo específico entre los especialistas del “front end” (usuarios) y los del “back end” (equipo de Data Warehouse); en este el diálogo debe ser continuo. Según sale a la luz un Data Warehouse, puede ser necesario realizar arreglos no planeados sobre el “back end” para soportar lo que la comunidad de usuarios pretende hacer con el sistema Data Warehouse. Esta alta integración requiere una comunicación continua entre los componentes del equipo de proyecto. Existen ciertas herramientas para que el usuario final realice procesamientos analíticos que necesitan de ciertas estructuras de datos específicas en el almacén y, en algunos casos, de estándares específicos en los nombres. Como resultado, a veces la naturaleza de las herramientas del “front end” nos llevan a diseñarlas de cierta forma.

2.3.3 FORMACIÓN DEL EQUIPO Y PERSONAS INVOLUCRADAS

La formación de un equipo depende de su situación particular que determina como se asignan estos papeles. Algunos requieren un equipo completo mientras que otros se pueden compartir entre una o dos personas.

Se proporciona una lista de referencia de los papeles a desempeñar en el Data Warehouse; éstas son áreas de responsabilidad que se necesitan cubrir cuando se construye un Data Warehouse:

Director de proyecto de Data Warehouse

Ésta persona que se elige debe tener un ingenio rápido y conocimiento sólido del negocio y su visión tecnológica. Mas importante aún, entiende por qué el negocio necesita un Data Warehouse y cómo vender este concepto a la dirección para obtener los fondos necesarios. El tiene la habilidad de hablar de la iniciativa de Data Warehouse con una perspectiva muy amplia. Debe pasar gran parte de su tiempo trabajando con la dirección para mantenerles al día del estado de éste y asegurarles que sigue las directrices del negocio. Generalmente, mantiene el control de gastos; es la clase de persona que tiene la perspectiva de la compañía y se asegura de que todos sus miembros la entiendan. Posee grandes cualidades de liderazgo.

Gestor de proyecto del Data Warehouse

Al contrario que el director de Data Warehouse, esta persona no necesita un ingenio rápido. Tiene profundos conocimientos de organización y capacidad de gestión. Debe consumir su tiempo gestionando la iniciativa y las directivas del

proyecto. Entiende la iniciativa del Data Warehouse desde el punto de vista del ratón en el laberinto; en cualquier caso, tener la capacidad de ver por encima del laberinto es importante.

Debe entender y supervisar todos los detalles. El plan del proyecto le pertenece y lo revisa a diario. Comprende cómo ser un gran negociador porque esta en constante resolución de conflictos y dando el orden de importancia de las tareas. Sabe qué es un camino y cómo gestionar el proyecto respecto de él. Trabaja constantemente con los líderes del equipo para valorar la solidez técnica de cada tarea.

Arquitecto del Data Warehouse

Es un técnico que vale para todo. Para alcanzar el éxito en este puesto, el entiende tanto el aspecto técnico (por ejemplo, familiaridad con el software y el hardware que se utiliza) como las necesidades del negocio. Debe, a partir de los descubrimientos hechos por el equipo de provisión de datos, encajarlos en esquema del Data Warehouse. Este esquema será, normalmente, el *diagrama entidad relación* que describen las tablas y las relaciones entre éstas.

El Data Warehouse contiene información de distintas áreas operacionales de la compañía y de algunos sistemas externos. Comprende todas estas entradas de datos y los entrelaza dentro de un único depósito integrado de datos. En una implantación exitosa, las distintas unidades de negocio tienen acceso a toda la información que necesitan para ser competitivos.

Si entiende las necesidades del negocio y las implementa en un diseño de base de datos funcional tiene la clave del éxito. No hay nada que sustituya a la

experiencia en el mundo real, porque esto es precisamente lo que le enseña cuándo seguir las reglas y cuando romperlas. Lo importante es que tenga la capacidad de escapar de los detalles y pueda tomar una visión más amplia.

Administrador de la base de datos

Una descripción perfecta de esta persona es aquella del rey benevolente. Para alcanzar el éxito en este trabajo, debe reinar sobre su base de datos. Rodea el castillo de la base de datos con un foso, bajando el puente sólo a aquellos que ganen el derecho a usarla. Sólo los amigos –nunca enemigos– pueden entrar dentro de ella. Como un rey o reina benevolente, es el responsable de la seguridad del castillo. Esto significa que se toman las medidas apropiadas y se establecen procedimientos para asegurar la base de datos y que se realizan las copias de seguridad.

Toma medidas de la respuesta de la base de datos y acciones correctivas para asegurar que el tiempo de respuesta es adecuado.

Este es un papel crítico que debe presentarse al principio del proceso. Muchas organizaciones cometen el error de no mantener un administrador de base de datos a jornada completa durante el proceso. Debe ser una extensión normal del equipo desde el principio para una efectividad completa.

Administrador del sistema

El administrador de sistema (SA) debe entender las entradas y las salidas del entorno de Data Warehouse elegido. Esta persona / equipo es responsable de la administración de los sistemas informáticos de igual forma que el

administrador de base de datos lo es de la administración de éstas. El SA es el responsable de las copias de seguridad del sistema, instalación del nuevo hardware, mejor del hardware, software de sistema, creación de un firewall (cortafuegos) y otras actividades del día a día que tienen que ver con los sistemas informáticos. Un firewall es un conjunto de programas de computador que restringen el acceso por red al Data Warehouse. Piense cuánto tiempo de proyecto se puede perder si se destruye el entorno de desarrollo.

Especialista en migración de datos (transferir los datos desde el sistema operacional o heredado)

Este trabajo es una piedra angular dentro del equipo. Si no se puede extraer los datos del sistema heredado, el proyecto de Data Warehouse falla. Es el responsable del desarrollo y adquisición de software que permite a la organización llevar los datos desde los sistemas heredados y las fuentes de datos externas a un área organizativa donde se realiza el esfuerzo de transformación. Ésta es una zona donde permanecen algunos datos antes de que se introduzcan dentro del repositorio de Data Warehouse.

Debe entender las diferencias entre el sistema heredado y el Data Warehouse. Su papel es transferir el conocimiento y los datos del sistema heredado a un formato en el que el equipo de transformación de datos pueda trabajar. El objetivo de la *migración de datos* es poner estos datos en un nuevo formato sin comprometer la integridad de los mismos y manteniendo todas las reglas de negocio. Ésta es una tarea tan crítica y con tanta complejidad que es útil separarla del esfuerzo de transformación de datos.

Si esta buscando un especialista en migración de datos, debe buscar en su organización a un veterano que posea un profundo conocimiento de trabajo con los sistemas heredados y un fuerte deseo de aprender nuevas tecnologías. Ésta es la típica persona de la organización que ha escogido un camino técnico en contra de uno de gestión. En este papel, la habilidad técnica es lo fundamental.

Es necesario que escriba o desarrolle herramientas internas para facilitar la migración porque ésta es un área en evolución dentro de la construcción del Data Warehouse. Debe estar abierto a soluciones de terceros.

Especialista en la transformación y filtrado de datos

La gente que ocupa este puesto se dedica a la migración de los datos a las tablas apropiadas junto con sus atributos asociados. Muchas veces, esta tarea incluye limpiar los datos y lograr que todos los códigos estén estandarizados. Un registro del sistema heredado puede dividirse en varias filas de distintas tablas, las cuales se pueden complicar rápidamente. Lo que parece una tarea sencilla no lo es. Éste esfuerzo consumirá una gran parte del proyecto. Básicamente, la tarea se parece a convertir una manzana en una naranja o a encajar una estaca redonda en un agujero cuadrado.

En resumen, se encarga de la transformación de hecho de los datos al nuevo diseño de base de datos. Así es cómo el antiguo sistema se transforma en uno nuevo.

Especialista en provisión de datos

Es quien va a reuniones con distintos jefes de departamento y con usuarios finales clave para definir las necesidades de negocio de la organización; después, comunican estas necesidades al resto del equipo. Para lograrlo, tiene gran facilidad de trato con la gente. Es capaz de extraer información de gente que no son técnicos y transmitir la información pertinente a los equipos técnicos. La información puede tomar distintas formas, desde prácticas reales del negocio a las facturas que se están utilizando. Posee suficiente inteligencia política para identificar a la gente clave en toma de decisión y a los usuarios clave. En resumen, capacitado para determinar que información debe estar disponible para que el Data Warehouse cubra las diferentes necesidades del negocio.

Líder de desarrollo de data mart

Un data mart es, por definición, un subconjunto del Data Warehouse. Las habilidades necesarias del líder de desarrollo de data mart se parecen mucho a las del especialista en provisión de datos. La única diferencia es que éste concentra su esfuerzo en un área determinada del negocio. Las habilidades y responsabilidades son las mismas que las del especialista en provisión de datos.

Especialista en pruebas y garantía de calidad

La calidad es algo que se aprende a la vieja usanza: trabajando en ello. Debe haber alguien que se concentre en la calidad. Si la cultura del equipo del Data Warehouse tiene fuertes raíces en la calidad, el producto final es de calidad.

Una forma de obtenerlo es teniendo un proceso de pruebas minucioso. Se establece en función de “guardián”, que es el responsable de la calidad.

Se involucra con los implicados para asegurar que todo el mundo trabaja en colaboración para crear un producto con un acabado perfecto. Normalmente, esta función es responsable de crear y aprobar los planes de pruebas. Este especialista es el responsable de revisar los resultados de las pruebas; esto puede significar asegurarse de que el código compila sin errores antes de llevarlo a producción o comparar los resultados del sistema heredado con los del Data Warehouse.

Especialista en infraestructura

Es el que se encarga de arreglar todos los detalles. Se asegura de que cada cosa está en su sitio. ¿Tienen los nuevos usuarios el hardware apropiado? ¿Tienen abiertas las cuentas? ¿Están definidos todos los papeles de acceso a la base de datos? Como puede verse, un Data Warehouse con éxito tiene cientos de detalles intrincados, cada uno de los cuales tiene que estar atendido para tener un proyecto con éxito. Es la persona responsable de todos los detalles más específicos.

Es un buen coordinador técnico para tener éxito. Debe ser capaz de desarrollar un plan de implantación y seguir su progreso a lo largo de la organización. sabe cuando hacer sonar el timbre de alarma y cuando actuar con cautela. Se puede hacer el mejor sistema del mundo, pero si los usuarios no pueden acceder a él, el trabajo habrá sido en vano.

Usuarios clave (especialista en el sistema operacional o heredado)

Nadie entiende la aplicación mejor que los usuarios. Ellos son los clientes finales. Si quiere tener éxito, tenga a los usuarios contentos y haga que se involucren. Tienen mucha mejor comprensión del negocio que el equipo técnico. Busque a aquellos usuarios que son los típicos apasionados por la tecnología y vayan a trabajar para usted adoptando rápidamente el nuevo sistema. Éste debe ser usado para ser aceptado. Cuanto antes se logre que los usuarios adopten y usen el nuevo sistema, pronto es aceptado el Data Warehouse.

Estos usuarios clave, normalmente, ayudan al equipo a saltar el abismo de conocimiento entre el antiguo sistema y el nuevo. Se les necesita para entender los sistemas actuales.

Entrenador

Un gran error por parte de muchas organizaciones es dar la tecnología a los usuarios sin un entrenamiento formal. Si quiere tener éxito, debe enseñar a los usuarios a cómo usar el Data Warehouse. Los mejores entrenadores son los típicos usuarios clave de dentro del negocio. Se recomienda que tenga un entrenador que desarrolle un buen material de entrenamiento y se lo dé a los usuarios clave para que enseñen en su departamento. En otras palabras, entrenar al entrenador. De esta forma, los usuarios clave refuerzan su conocimiento y mejoran su prestigio dentro del departamento. Es bueno que refuercen su base de poder porque están de nuestro lado. Los usuarios clave conocen a sus iguales mejor que nadie. Cuanto mas difundan las buenas noticias, mejor.

Escritor técnico

¿Que pasa si un desarrollador sufre un accidente camino al trabajo?
¿Existe documentación necesaria para proteger los intereses de la compañía?
Muchas veces, la respuesta a esta pregunta es "No". La mayoría de los desarrolladores no quieren escribir documentación técnica. Este es un hecho que debemos afrontar. Se necesita buena documentación para poner en conocimiento el trabajo que se está desarrollando. ¿Tiene más sentido tener un programador escribiendo documentación técnica o tener un escritor técnico haciendo la documentación? ¿Qué opción llevará a una mayor satisfacción en el trabajo y a un mejor producto? La respuesta es el escritor técnico; dejarlo escribir.

Su posición es frecuentemente descuidada. Un Data Warehouse es una aplicación compleja. Tener una buena documentación técnica estandarizada es útil. Ésta es una de las formas en que la comunidad de usuarios juzga el esfuerzo.

Se debe tener presente la situación en particular y las personalidades que se ven envueltas en el desarrollo del Data Warehouse. Es aconsejable que revise la situación desde dos perspectivas diferentes: la función y la persona. Debe decidir quien es el responsable de cada función y además determinar si la gente que se considera, tiene las habilidades apropiadas para las tareas necesarias.

2.3.4 GESTIÓN DEL PROYECTO

Es la aplicación de los conocimientos, habilidades, herramientas y técnicas para alcanzar o exceder las necesidades y expectativas de los implicados.

Lo que realmente gusta es el uso del término "implicados". Como todas las corporaciones, hay fiduciario que se responsabiliza de nuestros accionistas, de la misma forma en que un gestor de proyecto tiene una responsabilidad fiduciaria ante todos aquellos individuos y organizaciones afectados por los planes de proyecto y por su ejecución. Así, los "implicados" son aquellos grupos o individuos afectados por el proyecto y todas sus actividades. Estos son tanto los clientes como la gente involucrada internamente en el proyecto. Así es, cualquiera que esté involucrado tiene su parte de interés en la realización satisfactoria del plan. Cuanto antes se de cuenta de esto, mas probabilidad de éxito tiene.

Si se quiere obtener un proyecto con éxito, tomar el tiempo suficiente para identificar a todos sus implicados y entender qué impacto tienen en él. Debe entender cuáles son sus necesidades y sus expectativas. Sin éstas nunca puede gestionarles o persuadirles. Si no se toma el tiempo suficiente, el proyecto falla. Cuanto mejor se entiendan a los clientes, mas posibilidades de éxito se tiene. Como todas las decisiones que se toman en la vida, hay contrapartidas. La exitosa gestión de un proyecto también puede tener contrapartidas.

En muchos aspectos, el gestor de proyecto debe ser un gran diplomático, porque algunas de las expectativas y necesidades de los clientes pueden entrar

en conflicto. Esto es habitual. El gestor de proyecto es una persona capaz de mirar el problema desde distintos puntos de vista y tomar finalmente una decisión que sea la que mejor cumpla las necesidades del negocio de la forma mas eficiente, tanto en costos como en recursos.

Otra dolorosa lección que se aprende con los años es que no existen buenas herramientas que proporcionan la gestión del proyecto. Muchas organizaciones hacen que sea la "Herramienta de gestión del proyecto" la que determine la gestión del proyecto con todas las ventajas y desventajas que ésta tenga. Esto no funciona. Lo que hacen estas empresas es enviar a alguien a que le enseñen a usar una herramienta, no a aprender gestión de proyectos, que son cosas muy distintas.

Lo que se recomienda, y de hecho se hace, es que en la empresa se proporcione a todos los miembros una idea de las mejores técnicas de gestión de proyectos, es decir, hacer cursos especiales para cada empleado de la compañía. Así, algunos recibirán un curso de terminología y otros uno que les enseñe técnicas importantes, como la comprensión del camino crítico.

Lo que no se hace es dejar a la gente colgada con la forma de usar la herramienta. Muchas veces, algunos fracasan al estar demasiado ocupados haciendo una herramienta de gestión de proyecto. En vez de esto, deben estar ocupados en gestionar a su equipo las expectativas de los clientes y la comunicación. Hacer esto supone tratar asuntos que no pueden convenir. Como empresa, se decide contratar a un experto en la herramienta de gestión de proyectos y dejar a los gestores de proyecto libres para gestionar los proyectos.

Esto nos lleva otra vez a la cuestión clave: no deje que ésta sea la que determine la gestión. Desarrolle personal interno que sea experto en la "herramienta de moda". Esto dejara las manos libre a los expertos para que hagan lo que mejor saben.

- Valorar a cada interesado en el proyecto y determinar sus necesidades.
- Fijar las expectativas adecuadas en los clientes.
- Estar alerta ante los problemas y saber afrontarlos.
- Tener tiempo para gestionar el proyecto y no dejar que éste lo gestione a él.

También debe hablarse del proyecto, que es la causa última de cualquier iniciativa de gestión. Es un esfuerzo temporal emprendido para crear un producto o servicio único. Una de las claves aquí es "esfuerzo temporal". Esto significa que ha de tener definido un principio y un fin. Si no tiene claro lo que se debe hacer para acabar un proyecto, no lo empiece. Esto le lleva a un desastre seguro. Además, si en cualquier momento se ve claramente que el producto final no puede ser entregado, éste debe ser cancelado.

Otra de las claves es la de crear un producto o servicio "único". Se debe tener una idea muy clara de lo que está construyendo y de por qué es único. No es necesario que se entienda por que el producto es único desde el primer día, pero, según avance el proceso de desarrollo del plan, se van descubriendo los atributos que le hacen único. Cuando entienda lo especial que tiene, se esta en

condiciones de determinar qué constituye un proyecto completo. Sólo con un fin bien definido tendrá un proyecto que pueda tener éxito.

2.3.5 DATA WAREHOUSE VRS. DATA MART

La duplicación en otro entorno de datos es un término que suele ser mal interpretado e incomprendido. Así es usado por los fabricantes del Sistema de gestión de Base de Datos (SGBD) en el sentido de simple réplica de los datos de un sistema operacional centralizado en sistemas distribuidos. En un contexto de Data Warehouse, el término duplicación se refiere a la creación de Data Marts locales o departamentales basados en subconjuntos de la información contenida en el Data Warehouse central o maestro.

Un Data Mart es una aplicación de Data Warehouse, construida rápidamente para soportar una línea de negocio simple. Éstos, 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 éste.

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 necesitan tanto la segmentación como algunos métodos adicionales de consolidación.

La primera aproximación a una arquitectura descentralizada de Data Mart, puede venir originada de una situación como la descrita en la ilustración 2-16 que se aplica a una empresa de distribución de artículos.

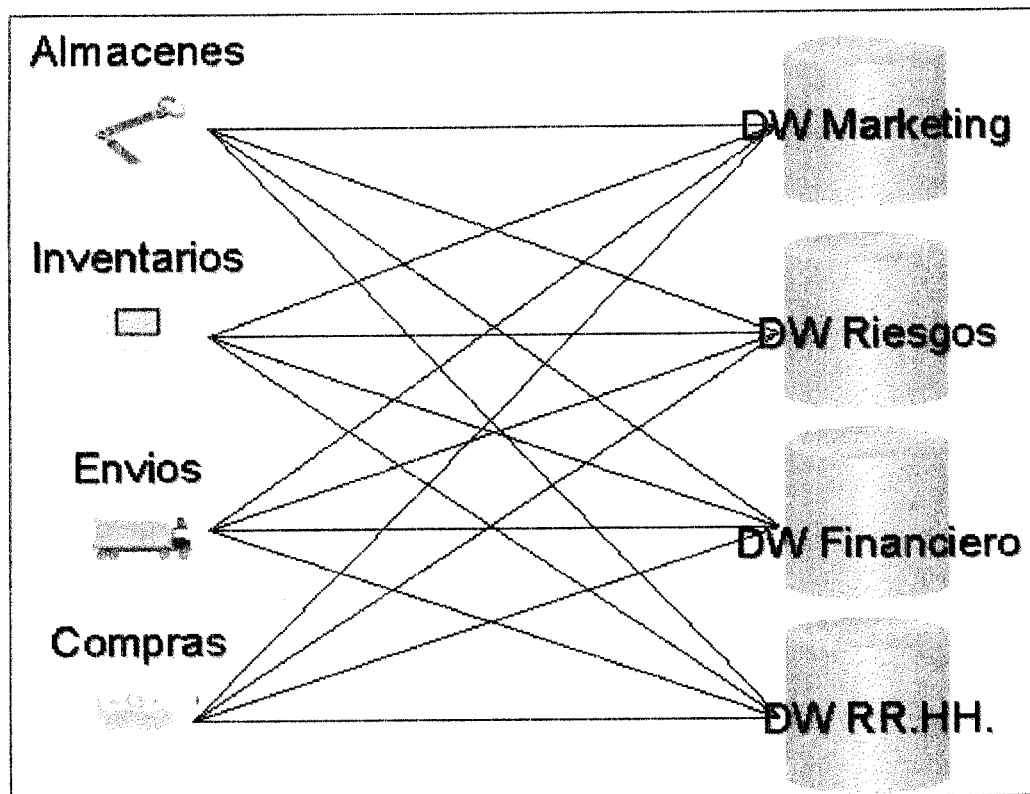


Ilustración 2-16

El departamento de Mercadeo, 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 los suyos. Mercadeo, comienza a usar otros datos que también usan los Data Marts de Riesgos y Financiero, y estos hacen lo propio.

Esto parece ser una decisión normal, puesto que las necesidades de información de todos crecen conforme el tiempo avanza. Cuando esta situación evoluciona, el esquema general de integración entre éstos pasa a ser como la ilustración 2-17

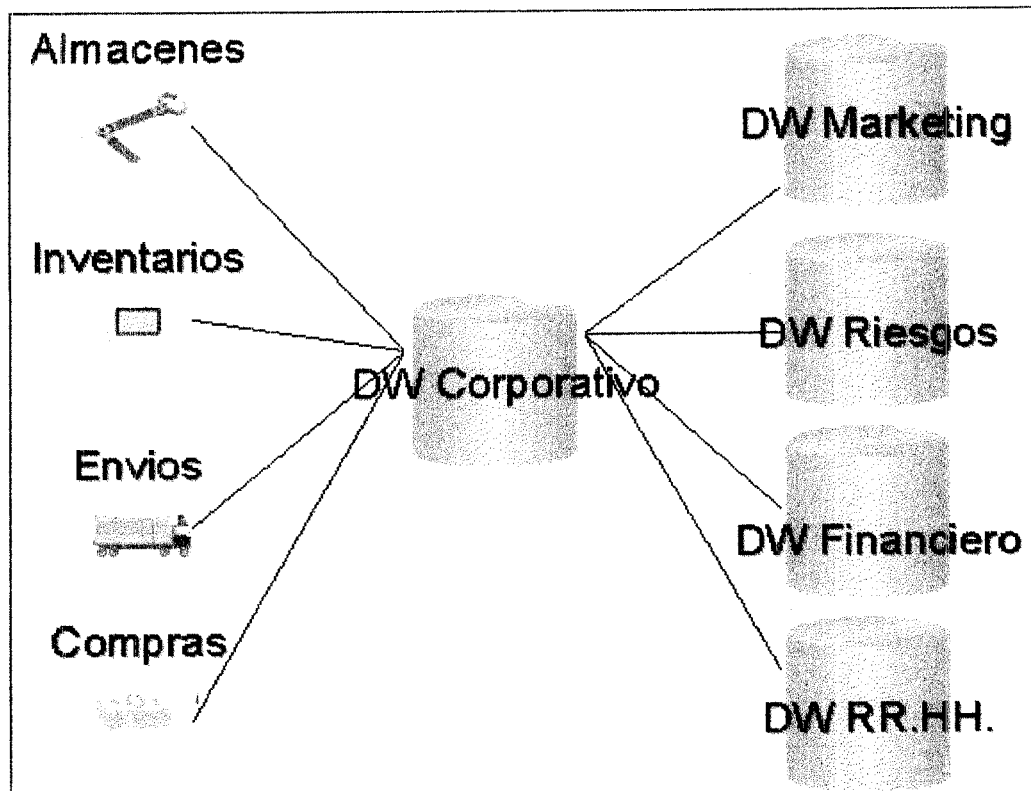


Ilustración 2-17

En esta situación, es fácil observar cómo éste esquema de integración de información de los Data Marts, pasa a convertirse en un rompecabezas en el que la gestión se ha complicado hasta convertir ésta ansia de información en un auténtico quebradero de cabeza. No obstante, lo que ha fallado no es la integración de Data Marts, sino la forma en que se integran.

En efecto, un enfoque más adecuado sería la coordinación de la gestión de información de todos los Data Marts en un Data Warehouse centralizado.

En esta situación se obtiene la información necesaria, ya previamente cargada y depurada en el Data Warehouse corporativo, simplificando el crecimiento de una base de conocimientos a nivel de toda la empresa.

Esta simplificación proviene de la centralización de las labores de gestión de los Data Marts, en el Data Warehouse corporativo, generando economías de escala en la gestión de los que están implicados.

2.3.6 COMPONENTES A TENER EN CUENTA A LA HORA DE CONSTRUIR UN DATA WAREHOUSE ORIENTADO A LA BANCA

2.3.6.1 HARDWARE

Un componente fundamental a la hora de poder contar con un Data Warehouse que responda a las necesidades analíticas avanzadas de los usuarios, es el poder contar con una infraestructura hardware que la soporte.

En este sentido son críticas, a la hora de evaluar uno u otro hardware, dos características principales:

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

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

En el mercado se desarrollan tecnologías basadas en el procesamiento paralelo, dan el soporte necesario a las necesidades de altas prestaciones y escalabilidad de los Data Warehouse. Estas tecnologías son de dos tipos:

- SMP (Symmetric multiprocessing, o Multiprocesadores Simétricos): Los sistemas tienen múltiples procesadores que comparten un único bus y una gran memoria, repartiéndose los procesos que genera el sistema, siendo el sistema operativo el que gestiona esta distribución de tareas. Estos sistemas se conocen como arquitecturas de "casi todo compartido". El aspecto más crítico es el grado de rendimiento relativo respecto al número de procesadores presentes, debido a su creciente no lineal.
 - MPP (Massively parallel processing, o Multiprocesadores Masivamente Paralelos): Es una tecnología que compite contra la SMP, en la que los sistemas suelen ser casi independientes comunicados por intercambiadores de alta velocidad que permiten
-

gestionarlos como un único sistema. Se conocen por ello como arquitecturas de "nada compartido". Su escalabilidad es mayor que la de los SMP.

Las tendencias de mercado indican que las arquitecturas SMP aportan normalmente suficientes características de escalabilidad, con una mayor oferta y un menor riesgo tecnológico. Sin embargo, cuando las condiciones de escalabilidad sean extremas, se puede plantear la opción MPP.

No obstante, se producen avances significativos en arquitecturas SMP, que han logrado máquinas con un crecimiento lineal de rendimiento hasta un número de 64 procesadores.

Se recomienda visitar la dirección de Internet: <http://www.tpc.org/> en donde la Transaction Processing Council (de la que son miembros ALR, Amdahl, Bull, Compaq, Data General, Dell, Digital, Fujitsu, HP, IBM, Intergraph, NCR, Siemens-Nixdorf, Sun o Unisys), realiza una comparativa entre las máquinas de sus miembros, proporcionando para diferentes modelos y diferentes configuraciones de Sistemas Operativos y Software de Base de Datos, un análisis de rendimiento (throughput), y un resumen de características (precio, número de procesadores, arquitectura y futuras versiones y fecha de disponibilidad).

2.3.6.2 SOFTWARE DE ALMACENAMIENTO (SGBD)

Conocido comunmente como Sistema de Gestión de Bases de Datos (SGBD), es el conjunto de programas que permiten definir, manipular y

utilizar la información que contienen las bases de datos (BD), realizar todas las tareas de administración necesarias para mantenerlas operativas, mantener su integridad, confidencialidad y seguridad. Una BD nunca se accede o manipula directamente sino a través de éste. Se puede considerar como la interfaz entre el usuario y la BD.

El funcionamiento del SGBD está muy interrelacionado con el del Sistema Operativo, especialmente con el sistema de comunicaciones. El SGBD utiliza las facilidades del sistema de comunicaciones para recibir las peticiones del usuario (que puede estar utilizando un terminal físicamente remoto) y para devolverle los resultados.

Como se comenta, el sistema que gestione el almacenamiento de la información, es otro elemento clave en un Data Warehouse. Independientemente de que la información almacenada en éste se pueda analizar mediante visualización multidimensional, el SGBD puede estar realizado utilizando tecnología de Bases de Datos Relacionales o Multidimensionales.

Las Bases de Datos Relacionales, se han popularizado en los sistemas operacionales, pero se han visto incapaces de enfrentarse a las necesidades de información de los entornos Data Warehouse. Por ello, las necesidades de información suelen atender a consultas multidimensionales, parece que las Bases de Datos Multidimensionales, parten con ventaja. En este sentido son de aplicación los comentarios hechos en el tema de

hardware, por requerimientos de prestaciones, escalabilidad y consolidación tecnológica.

Al igual que en el hardware, nuevos diseños de las bases de datos relacionales, las bases de datos post-relacionales, abren un mayor abanico de elección. Se basan en la combinación dinámica de los modelos de datos que están o no en la primera forma normal. Desde el punto de vista de las operaciones de tipo relacional, una tabla que contiene a otra anidada en ella, se visualiza como dos tablas unidas y cada una accesible como una entidad lógica, separada a través de un proceso llamado normalización dinámica. Este modelo es de una estructura tridimensional, es decir, los campos o grupos de éstos pueden aparecer varias veces, una vez, o nunca, sin limitación y sin necesidad de definición. Reduce el número de tablas y elimina la duplicación de datos. El modelo relacional tradicional (1NF) es un subgrupo del modelo post-relacional. Estas bases de datos post-relacionales, parten de una tecnología consolidada y dan respuesta al agotamiento de las posibilidades de los sistemas de gestión de bases de datos relacionales, ofreciendo las mismas prestaciones aunque implantadas en una arquitectura diseñada de forma más eficiente.

Esto se consigue instaurando relaciones lógicas en vez de físicas, lo que hace que ya no sea necesario destinar más hardware a una solución para conseguir la ejecución de las funciones requeridas. El resultado es que la misma aplicación implantada en una BD post-relacional requiere menos hardware, puede dar servicio a un mayor número de usuarios y utilizar

mecanismos intensivos de acceso a los datos más complejos. Asimismo, esta tecnología permite combinar las ventajas de las bases de datos jerárquicas y las relacionales con un coste más reducido. Ambos sistemas aportan como ventaja que no resulta necesario disponer de servidores omnipotentes, sin que puede partirse de un nivel de hardware modesto y ampliarlo a medida que crecen las necesidades de información de la compañía y el uso efectivo del sistema.

2.3.6.3 SOFTWARE DE EXTRACCIÓN Y MANIPULACIÓN DE DATOS

Se debe analizar un componente esencial a la hora de implantar un Data Warehouse, la extracción y manipulación. Para esta labor, que entra dentro del ámbito de los profesionales de tecnologías de la información, es crítico el poder contar con herramientas que permitan controlar y automatizar los continuos procesos y necesidades de actualización.

Estas herramientas deben proporcionar las siguientes funcionalidades:

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

para la construcción del Data Warehouse corporativo y de los Data Marts.

- Uso de la arquitectura de meta datos, facilitando la definición de los objetos de negocio y las reglas de consolidación.
- Acceso a una gran variedad de fuentes de datos diferentes.
- Manejo de excepciones.
- Planificación, bitácoras, interfaces a calendarizadores de terceros.
- Interfaz independiente de hardware.
- Soporte en la explotación del Data Warehouse.

A veces, no se suele prestar la suficiente atención a esta fase de la gestión del Data Warehouse, aun cuando supone una gran parte del esfuerzo en la construcción de éste. Existen multitud de herramientas disponibles en el mercado que automatizan parte del trabajo, por lo que debe de estudiarse cual es la que se adecua mejor.

2.3.6.4 HERRAMIENTAS MIDDLEWARE

Como herramientas de soporte a la fase de gestión de un Data Warehouse, se analizan a continuación dos tipos de herramientas:

- Por un lado herramientas Middleware (es un conjunto de componentes que facilitan tareas comunes a las aplicaciones que proveen autenticación, localización, acceso a recursos y movilidad), que suplen conectividad entre entornos diferentes, para ayudar en la gestión del Data Warehouse.
-

- Por otro, analizadores y aceleradores de consultas, que permitan optimizar tiempos de respuestas en las necesidades analíticas, o de carga de los diferentes datos desde los sistemas operacionales hasta el Data Warehouse.

Las herramientas Middleware deben ser escalables siendo capaces de crecer conforme lo hace el Data Warehouse, sin problemas de volúmenes. También deben ser flexibles y robustas, sin olvidarse de proporcionar un rendimiento adecuado. Tienen que estar abiertas a todo tipo de entornos de almacenamiento de datos, tanto mediante estándares de facto (OLE, ODBC, OLEDB.), como a los tipos de mercado más populares (DB2, Access, Oracle, Sybase, SQLServer.). La conectividad, al menos en estándares de transporte (SNA LU6.2, DECnet.) debe estar también asegurada.

Con el uso de estas herramientas de Middleware se logra:

- Maximizar los recursos ejecutando las aplicaciones en la plataforma más adecuada.
 - Integrar los datos y aplicaciones existentes en una plataforma distribuida.
 - Automatizar la distribución de datos y aplicaciones desde un sistema centralizado.
 - Reducir tráfico en la red, balanceando los niveles de cliente servidor (mas ó menos datos y proceso en local).
 - Explotar las capacidades de sistemas remotos sin tener que aprender múltiples entornos operativos.
-

- Asegurar la escalabilidad del sistema.
- Desarrollar aplicaciones en local y explotarlas en el servidor.

Los analizadores y aceleradores de consultas trabajan volcando sobre un fichero de bitácora las consultas ejecutadas y datos asociados a las mismas (tiempo de respuesta, tablas accedidas, método de acceso.). Ésta se analiza, automáticamente o mediante la supervisión del administrador de datos, para mejorar los tiempos de accesos.

Estos sistemas de monitoreo se pueden implantar en un entorno separado de pruebas, o en el entorno real. Si se ejecutan sobre un entorno de pruebas, el rendimiento del entorno real no se ve afectado. Sin embargo, no es posible optimizar los esfuerzos, puesto que los análisis efectuados pueden realizarse sobre consultas no críticas o no frecuentemente realizadas por los usuarios.

El implantar un sistema analizador de consultas, en el entorno real tiene además una serie de ventajas que permiten:

- Monitorear los tiempos de respuesta del entorno real.
 - Implantar mecanismos de optimización de las consultas, reduciendo la carga del sistema.
 - Imputar costos a los usuarios por el coste del Data Warehouse.
 - Implantar mecanismos de bloqueo para las consultas que vayan a implicar un tiempo de respuesta excesivo.
-

2.4 CONSTRUYENDO EL DATA WAREHOUSE ORIENTADO A LA BANCA

2.4.1 FASES DE IMPLANTACIÓN DEL DATA WAREHOUSE

Según Computer World: "Un Data Warehouse no se puede comprar, se tiene que construir". (<http://www.map.es/csi/silice/>) Tal como se ha mencionado con anterioridad, la construcción e implantación es un proceso evolutivo.

Este se tiene que apoyar en una metodología específica para este tipo de procesos, si bien es más importante que la elección de la mejor de las metodologías, el realizar un control para asegurar el seguimiento de la misma.

En las fases que se establecen 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 permite obtener resultados tangibles en un corto espacio de tiempo.

Se plantea aquí la metodología propuesta por SAS Institute: la "Metodología Rápida de Almacenamiento de Datos" (<http://www.map.es/csi/silice/>). La ilustración 2-18 (página 65) la muestra.

Ésta es iterativa, y se basa en el desarrollo incremental del proyecto de Data Warehouse dividido en cinco fases:

- Definición de los objetivos
 - Definición de los requerimientos de información
 - Diseño y modelado
 - Implantación
 - Revisión
-

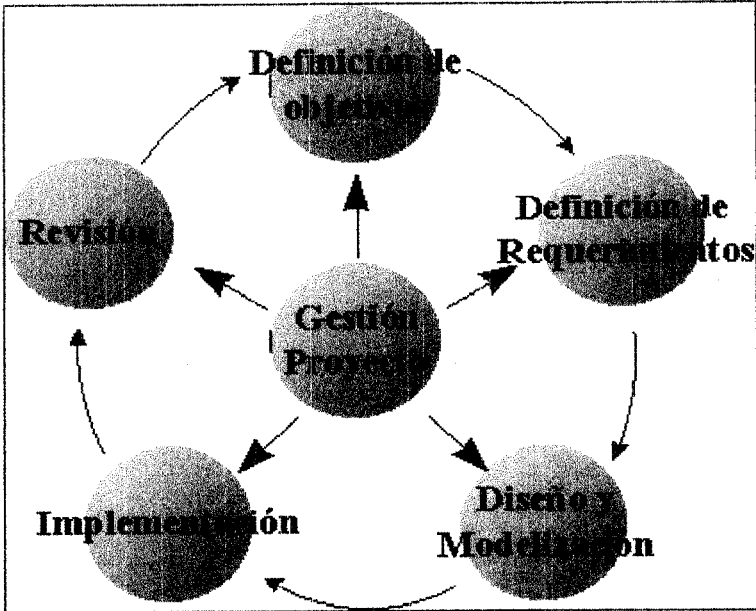


Ilustración 2-18

2.4.1.1 DEFINICIÓN DE LOS OBJETIVOS

Como todo proyecto, se debe definir sus objetivos y alcances que se pretenden conseguir. Esto demarca las etapas a las que se tienen que lograr en un determinado tiempo con la utilización de los recursos asignados para las tareas establecidas.

2.4.1.2 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, es analizar las necesidades y hacer comprender las ventajas que este sistema puede reportar.

2.4.1.3 DISEÑO Y MODELADO

Los requerimientos de información identificados durante la anterior fase proporcionan las bases para realizar el diseño y el modelado del Data Warehouse.

Se identifica las fuentes de los datos (sistema operacional, fuentes externas) y las transformaciones necesarias para, a partir de dichas fuentes, obtener el modelo lógico de datos del Data Warehouse. Este modelo esta formado por entidades y relaciones que permiten resolver las necesidades de negocio de la organización.

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

La mayor parte estas definiciones de los datos están almacenadas en los meta datos y forman parte del mismo.

2.4.1.4 IMPLANTACIÓN

Ésta fase 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
 - On-line analytical processing (OLAP: procesamiento analítico en línea)
 - Executive Information System (EIS: Información de gestión)
 - Decision Support Systems (DSS: Sistemas de soporte de decisiones)
 - Visualización de la información
-

- Data Mining ó Minería de Datos.

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

Con la finalización de esta fase se obtiene un Data Warehouse disponible para su uso por parte de los usuarios finales y el departamento de informática.

2.4.1.5 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, deber 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 son los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.

2.4.1.6 DISEÑO DE LA ESTRUCTURA DE CURSOS DE FORMACIÓN

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

2.4.2 ESTRATEGIAS DE IMPLANTACIÓN

Algunas consideraciones que se recomiendan deben seguirse a la hora de abordar un proyecto de este tipo son:

- *"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.

2.4.3 TÉCNICAS DE EXPLOTACION DE LA INFORMACIÓN

Dentro del esquema de Gestión y Explotación del Data Warehouse que se muestra en la ilustración 2-19, se pasa a detallar las posibilidades que nos ofrece esta última fase.

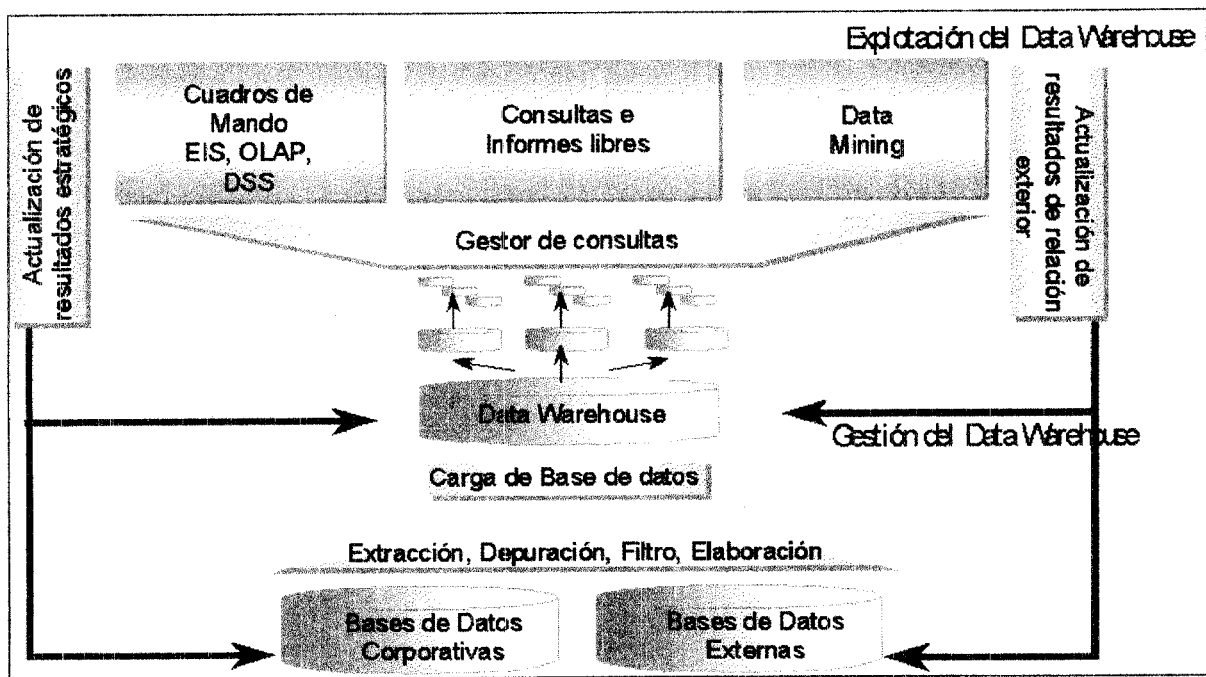


Ilustración 2-19

En esta fase se examina:

- El uso que se puede realizar de las utilidades OLAP del Data Warehouse para análisis multidimensionales,
- Las facilidades de obtención de información mediante consultas e informes libre, y el uso de técnicas de Data Mining que nos permitan descubrir "información oculta" en los datos mediante el uso de técnicas estadísticas.

2.4.3.1 OLAP, ROLAP, MOLAP

La explotación del Data Warehouse mediante información de gestión, se fundamenta básicamente en los niveles agrupados o calculados de información.

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

Éstas se estructuran a su vez en distintos niveles de detalle (por ejemplo, la dimensión geográfica puede constar de los niveles nacional, departamental, agencias locales y agencia central).

Este tipo de sistemas ha existido desde hace tiempo, en el mundo de la informática bajo distintas denominaciones: cuadros de mando, Management Information systems (MIS: Sistemas de información Gerencial), Executive Information systems (EIS: Sistemas de información de Gestión).

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.

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 deben:

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

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

Normalmente ésto se ve reflejado en la visualización de la estructura multidimensional, en unos campos de elección que nos permitan escoger el nivel de agregación (jerarquía) de la dimensión, y/o de un dato en concreto, la visualización de los atributos del sujeto, frente a unas dimensiones en modo tabla, permitiendo con ello realizar, entre otras las siguientes acciones:

- Rotar (*Swap*): alterar las filas por columnas (permutar dos dimensiones de análisis)
 - Bajar (*Down*): bajar el nivel de visualización en las filas a una jerarquía inferior
-

- Detallar (*Drilldown*): informar para una fila en concreto, de datos a un nivel inferior
- Expandir (*Expand*): igual al anterior pero sin perder la información a nivel superior para éste y el resto de los valores
- Colapsar (*Collapse*): operación inversa de la anterior.

Para ampliar sobre exploraciones en análisis OLAP, se recomienda visitar el sitio: <http://www.kenan.com/> en donde se describen en torno a 50 términos relacionados con las posibilidades de navegación que permiten este tipo de análisis.

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

2.4.3.1.1 Sistemas MOLAP

Ésta arquitectura usa unas Bases de Datos Multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos en múltiples dimensiones. Por el contrario, la arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales

Usa una base de datos propietaria multidimensional, en la que la información se almacena en muchas dimensiones, para ser visualizada de la misma forma.

Éste 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 por lotes. Al tener el dato elemental en la Base de Datos multidimensional (MDDDB), se realizan una serie de cálculos para los datos agregados, a través de las dimensiones de negocio, rellenando la estructura MDDDB.

Luego 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 completa, la MDDDB está lista para su uso. Los usuarios solicitan informes a través del interfaz, y la lógica de aplicación de la MDDDB obtiene el dato. La ilustración 2-20 (página 75) nos muestra gráficamente este proceso.

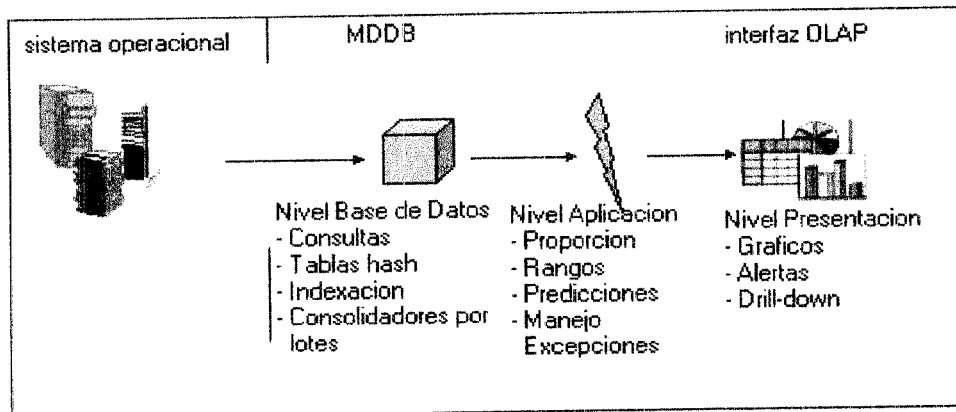


Ilustración 2-20

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos precompilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de encontrar proporciones de datos que no se precálculan y almacenan previamente.

2.4.3.1.2 Sistemas ROLAP

Ésta arquitectura, 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.

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 para agregar el dato, si así es requerido por éste.

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 devolverlos a los usuarios. La ilustración 2-21 muestra gráficamente el proceso.

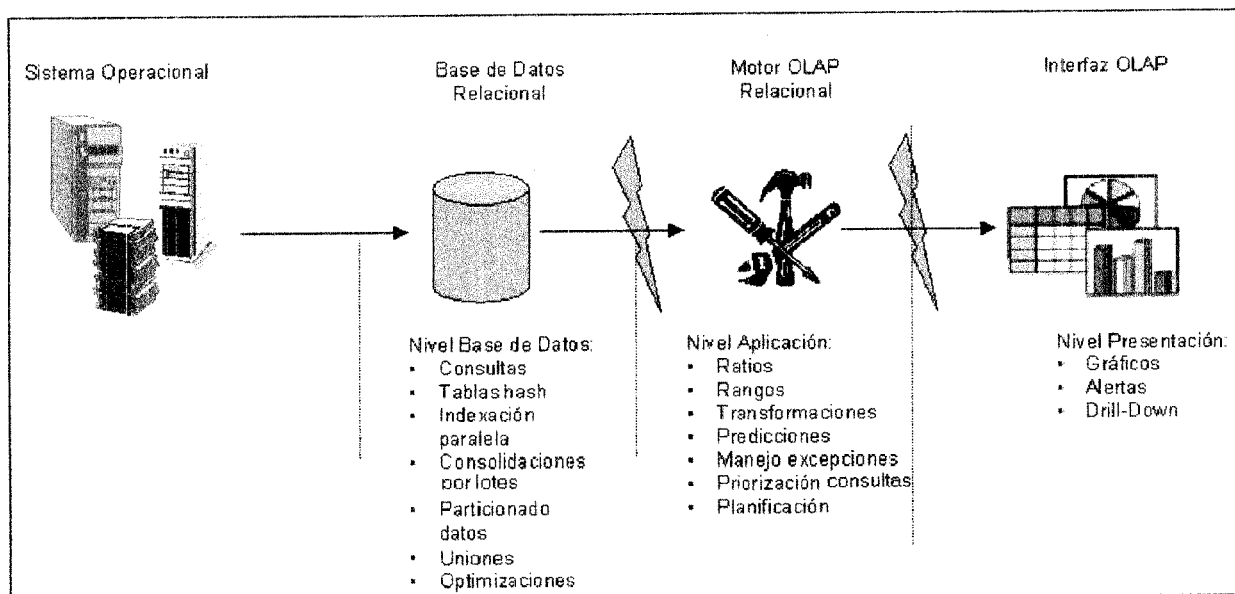


Ilustración 2-21

Ésta arquitectura 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 uniones múltiples.

2.4.3.1.3 ROLAP vrs. MOLAP (Comparativa)

Cuando se comparan las dos arquitecturas, se pueden realizar las siguientes observaciones:

- El ROLAP delega la negociación entre tiempo de respuesta y el proceso por lotes 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 por lotes.
 - 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 por lotes.
 - 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 Gigabytes)

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.

2.4.3.2 CONSULTAS Y REPORTEES

Las consultas o informes libres (queries) trabajan tanto sobre el detalle como sobre las agregaciones de la información.

Realizar este tipo de explotación en un Data Warehouse supone una optimización del tradicional entorno de informes (reporting), dado que éste mantiene una estructura y una tecnología mucho más apropiada para este tipo de solicitudes.

Los sistemas de consultas y reportes (Query & Reporting), no basados en almacenes de datos se caracterizan por la complejidad de las consultas, los altísimos tiempos de respuesta y la interferencia con otros procesos informáticos que compartan su entorno.

La explotación del Data Warehouse mediante "Query & Reporting" debe permitir un aumento de la flexibilidad de acceso, proporcional a la experiencia y formación del usuario. A este respecto, se recomienda el mantenimiento de al menos tres niveles de dificultad:

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

Hay una extensa gama de herramientas en el mercado para cumplir esta funcionalidad sobre entornos de tipo Data Warehouse, por lo que se puede elegir el software más adecuado para cada problemática empresarial concreta.

2.4.3.3 DATA MINING O MINERÍA DE DATOS

2.4.3.3.1 Introducción

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

Constituye por tanto una de las vías clave de explotación del Data Warehouse, dado que es este su entorno natural de trabajo.

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

2.4.3.3.2 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, A continuación se describen estas técnicas:

Análisis estadístico:

Utilizando las siguientes herramientas:

- *ANOVA*: o Análisis de la Varianza, contrasta si existen diferencias significativas entre las medidas de una o más variables continuas en grupo de población distintos.
 - *Regresión*: define la relación entre una o más variables y un conjunto de variables predictoras de las primeras.
 - *Ji cuadrado*: contrasta la hipótesis de independencia entre variables.
-

- *Componentes principales*: permite reducir el número de variables observadas a un menor número de variables artificiales, conservando la mayor parte de la información sobre la varianza de las variables.
- *Análisis cluster*: permite clasificar una población en un número determinado de grupos, en base a semejanzas y 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.

Métodos basados en árboles de decisión:

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

El árbol de decisión se construye partiendo el conjunto de datos en dos o más subconjuntos de observaciones a partir de los valores que toman las variables predictoras. Cada uno de estos subconjuntos vuelve después a

ser particionado utilizando el mismo algoritmo. Este proceso continúa hasta que no se encuentran diferencias significativas en la influencia de las variables de predicción de uno de estos grupos hacia el valor de la variable de respuesta.

La raíz del árbol es el conjunto de datos íntegro, los subconjuntos y los sub-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 la prueba de independencia de la Chi cuadrado sobre una tabla de contingencia.

Algoritmos genéticos:

Son métodos numéricos de optimización, en los que aquella variable o variables que se pretenden optimizar junto con las de estudio constituyen un segmento de información. Aquellas configuraciones de las variables de análisis que obtengan mejores valores para la 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 éstas (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 tienen que salir, 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, 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.

2.4.3.3.3 Metodología de aplicación

Para utilizar estas técnicas de forma eficiente y ordenada es preciso aplicar una metodología estructurada, al proceso de Data Mining. A este respecto se propone la siguiente metodología, siempre adaptable a la situación de negocio particular a la que se aplique:

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 es aquella que teniendo un error asumible contenga el número mínimo de observaciones.

En el caso de que se recurra a un muestreo aleatorio, se debe 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 debe informar del 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 ésta que nos proporcione el grado de error.

Exploración

Una vez determinada la población que sirve para la obtención del modelo se debe determinar cuales son las variables explicativas que sirven como "inputs" (entradas) 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.

También se pueden emplear técnicas estadísticas que nos ayuden a poner de manifiesto relaciones entre variables. A este respecto resulta 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.).

Modelado

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.

2.4.3.4 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:

- **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 navegador de Internet), simplificando el proceso de búsqueda de la información.
 - **Disponibilidad:** la información es accesible en todo momento, independientemente de los sistemas operacionales.
 - **Bajos costos de desarrollo y mantenimiento,** debido a la estandarización de las aplicaciones de consultas basadas en Internet, independientemente del sistema operativo que soporte el navegador, y de la reducción de los costos de distribución de software en los puestos clientes.
 - **Protección de los datos,** debido al uso de tecnologías consolidadas de protección en entornos de red (firewalls o corta fuegos).
 - **Bajos costos de formación,** debido al uso de interfaces tipo Web.
 - **La interactividad de las aplicaciones en este entorno pueden tener varios niveles:**
 - **Publicación de datos:** las páginas distribuyen información obtenida del Data Warehouse, volcada en las páginas intranet y/o 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 navegador de Internet o haciendo uso de "*programas*" de Java.

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

- Usar el Servidor Internet como enrutador, y ejecutar la petición desde el cliente al servidor directamente.
 - Hacer uso del navegador para visualizar una página Internet residente en el servidor de Internet. Esta página contiene información que se actualiza en éste, desde el servidor de Data Warehouse, a petición del usuario haciendo uso de interfaces comunes de salida (CGI's).
 - El cliente puede lanzar su consulta directamente al servidor de Data Warehouse, con "*programas*" de Java, haciendo el servidor Internet únicamente de enrutador (router).
 - El cliente puede ejecutar la aplicación de Data Warehouse desde el navegador, pero con un "agregado" (plug-in), que hace que se tengan las mismas opciones que posee la aplicación.
 - Realizar una descarga masiva de datos con un protocolo de transferencia de ficheros (FTP), para su proceso en local.
-

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, o CGI. Solución parcial, mientras la velocidad de transferencia se incrementa día a día mediante nuevos algoritmos de compresión de datos o el uso de líneas de alta capacidad.

2.4.4 TIPOS DE APLICACIONES EN LAS QUE UTILIZAR TÉCNICAS DISPONIBLES SOBRE EL DATA WAREHOUSE ORIENTADO A LA BANCA

2.4.4.1 DATA WAREHOUSE Y SISTEMAS DE MERCADEO

La aplicación de tecnologías de Data Warehouse supone un nuevo enfoque de Mercadeo, haciendo uso del "Mercadeo de Base de Datos". En efecto, un sistema de "Marketing Warehouse" implica un mercadeo científico, analítico y experto, basado en el conocimiento exhaustivo de clientes, productos, canales y mercado.

Este conocimiento se deriva de la disposición de toda la información necesaria, tanto interna como externa, en un entorno de Data Warehouse, persiguiendo con toda esta información, la optimización de las variables controladas del "Marketing Mix" y el soporte a la predicción de las variables no controlables (mediante técnicas de Data Mining). Basándose en el

conocimiento exhaustivo de los clientes se consigue un tratamiento personalizado de los mismos tanto en el día a día (atención comercial) como en acciones de promoción específicas.

Las áreas en las que se puede aplicar las tecnologías de Data Warehouse a Mercadeo son, entre otras:

- Investigación Comercial
- Segmentación de mercados
- Identificación de necesidades no cubiertas y generación de nuevos productos, o modificación de productos existentes
- Fijación de tasas de interés
- Definición de la estrategia de canales de comercialización y distribución
- Definición de la estrategia de promoción y atención al cliente
- Relación con el cliente:
- Programación, realización y seguimiento de acciones comerciales
- Lanzamiento de nuevos productos

2.4.4.2 DATA WAREHOUSE Y ANÁLISIS DE RIESGO FINANCIERO

El Data Warehouse aplicado al análisis de riesgos financieros ofrece capacidades avanzadas de desarrollo de aplicaciones para dar soporte a las diversas actividades de gestión de riesgos. Es posible desarrollar cualquier herramienta utilizando las funciones que incorpora la plataforma, gracias a la potencialidad estadística aplicada al riesgo de crédito.

Así se puede usar para llevar a cabo las siguientes funcionalidades:

- Para la gestión de la posición:
Determinación de la posición, Cálculo de sensibilidades, Análisis “what/if”, Simulaciones, Monitoreo de riesgos contra límites.
- Para la medición del riesgo:
Soporte metodología RiskMetrics (Metodología registrada de J.P. Morgan / Reuters), simulación de escenarios históricos, modelos de varianzas, modelos de valoración, análisis de rentabilidad, establecimiento y seguimiento de límites, desarrollo / modificación modelos, pruebas de tensión (stress testing).

El uso del Data Warehouse ofrece una gran flexibilidad para creación o modificación de modelos propios de valoración y medición de riesgos, tanto motivados por cambios en la regulación, como en avances en el modelado de estos instrumentos financieros.

Ello por cuanto se puede almacenar y poner a disposición información histórica de mercado y el uso de técnicas de Data Mining nos simplifica la implantación de cualquier método estadístico. Los métodos de previsión, se pueden realizar usando series históricas.

Pero la explotación de la información nos permite no solo la exploración de los datos para un conocimiento de la información histórica, sino también para examinar condiciones de normalidad de las que la mayoría de las metodologías de valoración del riesgo parten.

Además de implantar modelos ya existentes, se pueden acometer análisis con vistas a determinar modelos propios, basados en análisis de correlación para el estudio de la valoración del riesgo de carteras.

Todo ello en una plataforma avanzada de gestión de la información basada en la fácil visualización de la misma y de su análisis estadístico como soporte a metodologías estándar de facto, o a las particularidades de cada entorno.

2.4.4.3 DATA WAREHOUSE Y ANÁLISIS DE RIESGO DE CRÉDITO

La información relativa a clientes y su entorno se ha convertido en fuente de prevención de Riesgos de Crédito. En efecto, existe una tendencia general en todos los sectores a recoger, almacenar y analizar información crediticia como soporte a la toma de decisiones de Análisis de Riesgos de Crédito.

Los avances en la tecnología de Data Warehouse hacen posible la optimización de los sistemas de Análisis de Riesgo de Crédito:

Para la gestión del riesgo de crédito los sistemas operacionales han ofrecido:

- Sistemas de Información para Gerencia e informes de Soporte a la Decisión de Problemas estáticos y no abiertos a nuevas relaciones y orígenes de datos, situación en la que la incorporación de nuevas fuentes de información ha sido un problema en lugar de una ventaja.
-

- Exploraciones de datos e informes cerrados y estáticos.
- Análisis sin inclusión de consideraciones temporales lo que imposibilita el análisis del pasado y la previsión del futuro.
- Herramientas de "credit-scoring" no flexibles, construidas sobre algoritmos difícilmente modificables, no adaptados al entorno de la empresa, o exclusivamente basados en la experiencia personal no contrastada, con lo que los sistemas han ayudado a repetir los errores en vez de a corregirlos.

Pero estos sistemas tradicionales se enfrentan a una problemática difícil de resolver para acomodarse a las necesidades analíticas de los Sistemas de Análisis del Riesgo, necesidades que se pueden cubrir mediante el uso de tecnologías de Data Warehouse

Dentro de la *Prevención de créditos incobrables*, utilizando sistemas OLAP se puede obtener el grado interno de concentración de riesgos con el cliente, y almacenar la variedad de fuentes internas o externas de información disponibles sobre el mismo. Ello permite obtener sin dificultad la posición consolidada respecto al riesgo del cliente. El análisis se puede realizar asimismo por las diferentes características de la operación para la que se hace el análisis, en cuanto al plazo y la cuantía de la misma, la modalidad de crédito elegida, la finalidad de la operación o las garantías asociadas a la misma. Usando las mismas capacidades es fácil el establecer una segmentación ABC de la cartera de clientes potenciales o reales que nos optimicen el nivel de esfuerzo en el Análisis de Riesgos.

En el soporte al proceso de *Anticipación al Riesgo*, se puede dar un adecuado soporte a la correcta generación y consideración de señales de alerta, teniendo en cuenta las pautas y condicionantes diferenciados dependiendo del tipo de cliente y producto usando Data Mining

Para el caso del *Seguimiento del ciclo de créditos incobrables*, de nuevo el uso de sistemas OLAP, simplifican el análisis la diversidad de los diferentes parámetros que intervienen en el mismo, tales como la jerarquía de centros de recobro a contemplar, la diferente consideración dependiendo de la antigüedad de la deuda, del cliente o del importe adeudado. Un sistema de Data Mining puede aconsejar la mejor acción en caso de créditos incobrables, litigio, frente a los parámetros de importe, antigüedad, zona geográfica.

Estos sistemas hacen que el analista se dedique con más intensidad al análisis de la información, que es donde aporta su mayor valor añadido, que a la obtención de la misma. No obstante, estos sistemas deben de huir de las automatizaciones completas sin intervención del analista: es él que mejor sabe lo que quiere descubrir. "*La herramienta debe ser un medio y no un fin*".

2.4.4.4 DATA WAREHOUSE: OTRAS ÁREAS DE APLICACIÓN

Otras áreas de la empresa que pueden aplicar las soluciones que proporciona la tecnología Data Warehouse para mejorar gran parte de sus procesos actuales son:

- Control de Gestión:
Sistemas de Presupuestación, Análisis de Desviaciones, Reporting (EIS, MIS)
- Logística:
Mejora de la relación con proveedores, Racionalización de los procesos de control de inventarios, Optimización de los niveles de producción, Previsión de la demanda en infraestructura.
- Recursos Humanos:
Planificación de incorporaciones, Gestión de carreras profesionales, Asignación de recursos a proyectos alternativos.

2.4.5 SEGURIDAD DE ACCESO Y MANIPULACIÓN DE LA INFORMACIÓN EN EL DATA WAREHOUSE

Se deben tener ciertas consideraciones a contemplar en cuanto a seguridad de accesos y seguridad de datos (backup), puesto que si bien ésta (a 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 están validados. Debe contar con éstos y 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.

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:

- 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.
-
-

CAPITULO III

3. ASPECTOS TÉCNICOS EN EL PROCESO DE ADQUISICIÓN DE SISTEMAS DE DATA WAREHOUSE

Se debe dar la orientación suficiente al comprador para la preparación del conjunto de especificaciones que definirán los requisitos que han de cumplir la Creación y Explotación de un Data Warehouse.

Se realiza en primer lugar un análisis de las necesidades del comprador, a continuación se recogen los factores relevantes a tener en cuenta en el proceso de adquisición y, finalmente, se describe cómo deben ser planteadas las especificaciones técnico - funcionales para la elaboración del Pliego de Prescripciones Técnicas, qué normas, estándares y cláusulas pueden ser de aplicación, y cuál es el cuestionario técnico diseñado para normalizar las ofertas y facilitar su evaluación.

3.1 ANÁLISIS DE LAS NECESIDADES DEL COMPRADOR

En efecto, como punto de arranque de todo, es preciso "vender la idea" a los usuarios finales de un Data Warehouse. Esto es así, por ser bastante novedosa y sobre la que pueden surgir recelos de su efectividad. Estos se pueden eliminar comenzando por un pequeño módulo, del cual se valoren los beneficios posteriores, para iniciar progresivamente el desarrollo de nuevos, cada uno con un costo unitario cada vez más reducido, pero sin embargo con unos beneficios distribuidos cada vez mayores por poder cada vez incluir más información.

El simple hecho de realizar un informe de necesidades previas en el que se enumeren la situación de los datos entre los diversos sistemas operacionales, puede ser un hecho decisivo para emprender un proyecto de este tipo. Muchas veces la información se encuentra tan poco normalizada y existen tantas discrepancias entre estos sistemas, que el abordar un Data Warehouse en el que se limpien estos datos y se normalicen pueden aportar un valor intangible: *"la calidad y fiabilidad de la información"*.

La venta de esta idea no sólo se ha de realizar frente a la Dirección sino que es preciso realizarla a todos los niveles: a la Dirección, Gerencia e incluso al área de Desarrollo.

Tras ésto, comienza dos fases similares al análisis de requisitos del sistema: la definición de objetivos y requerimientos de información, en el que se analicen las necesidades del comprador.

3.1.1 DEFINICIÓN DE LOS OBJETIVOS

En esta fase se define el equipo de proyecto que debe estar compuesto por representantes del departamento informático y de los departamentos usuarios del Data Warehouse además de la figura del director del proyecto.

Se define el alcance del sistema y cuales son las funciones que el Data Warehouse realiza como suministrador de información de negocio estratégica para la empresa. Se estipula así mismo, los parámetros que permitan evaluar el éxito del proyecto.

3.1.2 DEFINICIÓN DE LOS REQUERIMIENTOS DE INFORMACIÓN

Durante esta fase se mantienen sucesivas entrevistas con los representantes del departamento usuario final y los representantes del departamento de informática. Se realiza el estudio de los sistemas de información existentes, que ayudan a comprender las carencias actuales y futuras que deben ser resueltas en el diseño del Data Warehouse

Asimismo, en esta fase el equipo de proyecto debe ser capaz de validar el proceso de entrevistas y reforzar la orientación de negocio del proyecto. Al finalizar esta fase se obtiene el documento de definición de requerimientos en el que se refleja no solo las necesidades de información de los usuarios, sino cual es la estrategia y arquitectura de implantación del Data Warehouse.

3.2 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 varía en función de cada caso particular, por lo que siempre es necesario identificar la importancia relativa de cada punto. Éstos son bastante similares a los contemplados en el tema de los SGBD, que se adaptan al caso particular de un Data Warehouse.

3.2.1 PRUEBAS EN CONDICIONES REALES

Tal y como sucedía en los SGBD, el rendimiento real de un Data Warehouse es muy difícil de predecir mediante procedimientos teóricos. Por ello, de igual forma que en el SGBD, si se va a instalar un Data Warehouse que contiene un gran volumen de datos o, si por cualquier otra razón, existen dudas sobre la capacidad del mismo de dar el procesamiento adecuado en las máquinas disponibles se debe exigir al proveedor una prueba anterior a su adquisición. Esta prueba debe realizarse en la propia instalación de destino.

La prueba se debe realizar en las condiciones más parecidas a las reales que se puedan conseguir. Para ello se debe cargar el Data Warehouse con un volumen de datos adecuado y se debe crear procesos de prueba similares a los más costosos de los que se van a desarrollar.

A diferencia de con los SGBD, 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 es 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.) implicados.

3.2.2 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, en este sentido y como factor común de un SGBD con un Data Warehouse, cabe reseñar que muchos problemas de rendimiento se deben más a un mal diseño del modelo de datos del Data Warehouse que a un problema de rendimiento de la herramienta en sí.

3.2.3 DIMENSIONAMIENTO DE LA PLATAFORMA DE INSTALACIÓN

De lo comentado en los dos puntos anteriores puede deducirse que existe la posibilidad de que sea necesario aumentar la capacidad de la máquina en la que se instale el Data Warehouse, o mejor aun, disponer de una dedicada al Data Warehouse.

Es necesario que el proveedor detalle cual de las dos versiones está ofreciendo para cada una de las licencias que se compran y si alguna de ellas

fuese una versión limitada, que especifique claramente cuales de las funcionalidades ofertadas no se encuentran presentes en la versión restringida.

3.2.4 CONDICIONES ECONÓMICAS Y DEL SOPORTE

Existen actualmente varios sistemas de cobro por el uso de Data Warehouse según el fabricante. Los más utilizados son facturar por:

- Cada máquina y/o tipo de máquina en la que se instale.
- Cada usuario que acceda al SGBD.
- Por tiempo de utilización (usualmente renovación anual).
- Por combinación de las anteriores.

Es imprescindible que el proveedor indique con toda claridad el método utilizado.

También debe estar explícito que, salvo indicación en contrario, todas las licencias son de versiones completas sin ninguna restricción respecto a las funcionalidades ofertadas.

También es conveniente pedir los precios de los productos adicionales que no se desee instalar en el momento pero que puedan ser interesantes en el futuro.

Un factor importante es la duración de la garantía, período de tiempo durante el que el proveedor proporciona soporte gratuito a sus productos y, también el precio del soporte en años sucesivos. Este último precio debe fijarse sobre variables presentes en el contrato no sobre futuros precios de lista del fabricante.

Otro factor que debe evaluarse es la calidad del soporte ofrecido. Este puede dividirse en un gran número de puntos cuya importancia varía en función de las necesidades del comprador. Entre ellos, se pueden citar:

- La inclusión o no de la instalación en el precio del producto.
- El tiempo máximo de entrega.
- La inclusión o no de soporte adicional gratuito como puede ser un cierto número de horas de formación.
- Capacitación y experiencia del personal que presta soporte técnico y consultoría.
- Calidad de la documentación, idioma en que está escrita, número de copias suministradas gratuitamente y precio de las copias adicionales.
- Capacidad técnica de la empresa y de la asistencia técnica que presta para lo que es recomendable pedir referencias a otros usuarios de la Administración de este tipo de productos.

3.3 TENDENCIAS TECNOLÓGICAS Y DE MERCADO

Se describe una recopilación de las principales tendencias observadas en el mercado, que fueron explicadas en el capítulo anterior.

Tendencias hacia herramientas especializados

El uso de herramientas de propósito general no satisface por completo las necesidades de un proyecto de Data Warehouse. Debe hacer énfasis en las herramientas de usuario final y tecnológicas acorde a las necesidades propias.

Webhousing

El uso de Internet como fuente de información hacia el exterior e interior (vía intranets), crece constantemente, y la integración de una herramienta de Data Warehouse con Internet.

Uso generalizado de Data Marts

Las peculiaridades de un proyecto Data Warehouse, y el enfoque progresivo de su construcción, hace que cada vez mas organizaciones realicen sus desarrollos mediante el uso de Data Marts integrados.

Uso de tecnología OLAP

Enfocado al análisis multidimensional de los datos.

CAPITULO IV

4. PRUEBAS DE VERIFICACIÓN Y CONTROL

Un Data Warehouse precisa de un conjunto de pasos desde su creación hasta su explotación final. En todos estos pasos se precisa de una herramienta que nos permita simplificar su ejecución. Debido a la amplia variedad de herramientas disponibles, y para asegurar la idoneidad de la herramienta, caben varias alternativas:

- **Evaluación distribuida:** En ella se evalúa cada herramienta en un paso concreto. Se ve cual es la que mejor se adapta a nuestras necesidades, y a continuación, examina la compatibilidad en cascada de todas ellas.
- **Evaluación centralizada:** En ella se evalúa el soporte que cada herramienta proporciona a todos los pasos, y a continuación se examina cual es la que globalmente se adecua de forma global a nuestras necesidades.

La elección de una u otra alternativa tiene sus ventajas e inconvenientes: la primera nos permite obtener la mejor herramienta (la de mejor relación cumplimiento expectativas/ precio), pero puede comprometer la cadena completa de uso de un Data Warehouse. Lo contrario ocurre en la segunda alternativa.

Es recomendable acudir a fuentes de información sobre los siguientes aspectos:

Vendedores de herramientas de usuario final:

- Herramientas de Consultas y Reportes (*Query and Reporting*)
- Bases de Datos OLAP y Multidimensionales
- Sistemas EIS (Executive Information Systems)
- Herramientas de Data Mining
- Recuperación de Documentos
- Sistemas de Información Geográfica (GIS)
- Herramientas de Análisis de Decisiones
- Estadísticas
- Modelado de Procesos
- Filtrado de Información
- Obtención de Informes
- Otras herramientas

Vendedores de Infraestructura Tecnológica:

- Extracción de Datos, Limpieza y Carga
 - Catalogación de la Información
 - Bases de Datos para Data Warehousing
 - Administración de Consultas y Almacenamiento
 - Modelado de Datos para Data Warehouse
 - Herramientas Middleware
 - Aceleradoras de Consultas y Carga
 - Otras utilidades de BD
 - Hardware
-

Vendedores por Función e Industria:

- Herramientas de Análisis Financiero
- Herramientas de Mercadeo y Análisis de Ventas
- Herramientas de Análisis de la cadena de suministro
- Herramientas para la Industria de la Salud
- Herramientas para la Industria Detallista
- Herramientas para el Sector Financiero
- Otras herramientas específicas

Empresas de Servicios:

- Consultoras Especialistas en DW
 - Consultoras Generales
 - Formación
-

CAPITULO V

5. POLITICAS NECESARIAS

El primer paso hacia la construcción de un Data Warehouse exitoso comienza con la realización. El camino puede estar en desorden, con tanta gente y con burocracia, que son obstáculos de la tecnología. Las batallas, la resistencia del usuario, y las luchas de fuerzas pueden ser tan críticas al éxito de un Data Warehouse como eligiendo el “back end” correcto de la base de datos o del diseño del esquema. Éstos pueden ser politizados por que ellos:

- Cruzan líneas de organización;
- Cambia ambos términos de la propiedad de los datos y acceso, posiblemente exponiendo la historia de la administración de datos en la organización; y
- Afecta las prácticas del trabajo de las comunidades de usuarios altamente individualistas y de gran alcance en la empresa.

Dos cuestiones claves capaces de encender una discusión política alrededor del desarrollo de un Data Warehouse son:

- La pérdida de control sobre el acceso a la información en bruto por sí mismo; y
- La amenaza que el Data Warehouse mueve la toma de decisiones dentro de la infraestructura tecnológica de la empresa. Ambas cuestiones pueden conducir a una indisposición de los trabajadores con conocimiento para participar en el diseño, la construcción, y el uso del Data Warehouse y del microanálisis sin

fin, del significado de calidad y realidad de los datos proporcionados por el almacén.

Algunas señales de peligro o de malestar político pueden ser estas:

- El proyecto es conducido por un sentido fuerte pero pobremente definido de urgencia, no claramente conectado a las necesidades del negocio o empresa y/o a los cambios en la viabilidad de la organización;
 - Demandas hechas por la tecnología de información o la gerencia de administración del negocio para una justificación de costo o eventos del negocio sin claras indicaciones de los objetivos que justifican esos eventos;
 - Carencia de un fuerte comité que vele por el proyecto;
 - Los patrocinadores de proyecto no incluyen ambas representaciones, tecnología de información y representación del negocio;
 - El propósito del proyecto se centra en datos más bien que en ingresos, la satisfacción de cliente, u otro objetivo de negocio;
 - La gente adentro y afuera del equipo de proyecto que describe el Data Warehouse sobre todo en términos de sus componentes tecnológicos;
 - Miembros del equipo del proyecto que no pueden explicar a sí mismos cómo el destacamiento y el uso del almacén resolverán materialmente las necesidades del negocio;
-

- Las organizaciones de usuarios finales no dedican constantemente tiempo, en el nivel individual, al análisis del almacén, al diseño, y a las actividades piloto;
- La gerencia de administración del negocio no toma un interés proactivo en el estado del proyecto, es decir, tiene que ser obligado para tomar un interés; y
- Los detractores y los escépticos en comités de usuarios finales se hacen cada vez más elocuentes mientras que progresa el proyecto.

Para evitar la politización de proyectos de Data Warehouse, se sugiere las contramedidas siguientes:

- Cambie su mentalidad. Un diseñador del almacén debe ser sociólogo, un comercializador, diplomático, y tecnológico, en ese orden. Pase más tiempo considerando sus comités y sus necesidades, cómo usted las comercializa a ellas, y cómo usted establece acuerdos y lazos de trabajo con ellos tan así que usted pasa dibujando modelos de los datos y diagramas de arquitectura de la tecnología.
 - Consiga comodidad con ser franco; particularmente con los propietarios de los datos. Prevea y diga que usted espera encontrar los datos sucios, contrarios, e incompletos. Ayude a los propietarios de los datos para que se sientan cómodos con esta expectativa como un "estado de la naturaleza" mas bien que
-

algunos que fallan de su parte. Ayúdeles a limpiarla. Precise a ellos que el proyecto de almacenamiento es una oportunidad ideal para sacar desde abajo su carga pesada para limpiar los datos.

- Haga su análisis sociológico primero. Conozca a todos sus electorados: sus datos y acceso; requisitos técnicos y del negocio; sus perfiles de riesgo / recompensa; qué necesitan dar para que el proyecto tenga éxito; y qué ganar si el proyecto es exitoso.
 - Desarrolle un plan interno de comercialización antes de hacer el primer diseño. Sepa a quién tiene que vender el proyecto, cuál es el valor para cada público objetivo, y cómo se sabe cuándo tiene adentro su compra.
 - Establezca acuerdos / contratos bilaterales claros con todos los propietarios de datos antes de que la fase del diseño del proyecto se termine. Cada acuerdo debe especificar quién consigue a quién, cuando, y bajo qué circunstancias.
 - Relance y reajuste regularmente las expectativas con cada comité, cara a cara.
 - Pase por lo menos 10 minutos de cada reunión del proyecto discutiendo el clima político que rodea el proyecto, cambios en los comités, estados de los esfuerzos de comercialización internos, y desmienta los últimos chismes y rumores.
 - Pase por lo menos 15 minutos de cada reunión revisando el trabajo en marcha en términos de los objetivos de negocio para el
-

proyecto, y de las necesidades de los usuarios del comité y de su ambiente de trabajo.

- Tenga un proceso formal para registrar, revisar, y rechazar o validar (con los cambios de costo y de tiempo conocidos formalmente) todos los cambios al alcance del proyecto.
 - Pase mucho tiempo en contacto con los mayores patrocinadores técnicos y del negocio del proyecto. Cerciórese de que sepan, cada semana: el estado del proyecto, las barricadas de organización, y qué es para el equipo de proyecto que los patrocinadores hagan sobre esas barricadas. Utilice a los patrocinadores para reducir, vía orden administrativa, la resistencia de la organización que el equipo de proyecto no pueda quitar por la vía diplomática.
-

CAPITULO VI

6. EL ÉXITO DEL DATA WAREHOUSE

El éxito final de un Data Warehouse depende de cómo y porqué el almacén fue construido (ej., el crecimiento efectuado). Por lo tanto, los resultados del almacenamiento de los datos pueden variar. Por ejemplo, con el desarrollo por fases, el rendimiento de la inversión para el almacén es difícil de determinar, como el énfasis se centra más en la compartición de datos, los costos totales, y la mitigación de las cuestiones de posesión de los datos.

Dos medidas del éxito de un Data Warehouse son su uso y/o el rendimiento de la inversión. Un Data Warehouse alcanza su valor más grande cuando se utiliza en exceso. Se dice que se conoce que se es acertado cuando los eufóricos usuarios exigen aún más tipos de datos y tener acceso a más detalle. Más allá de los usuarios que exigen más acceso y características, la empresa debe gozar de beneficios materiales, ej., las reducciones de costos operativos debido a cortes en los ciclos de informes, seguimiento al cliente.

Los beneficios materiales se pueden medir como un rendimiento de la inversión tradicional. Sin importar la medida usada, la terminación exitosa y el uso de un Data Warehouse depende más de la cultura (es decir, reactivo o proactivo) de la empresa que su profesionalismo técnico.

BIBLIOGRAFIA

Abbey Michael-Corey Michael

ORACLE DATA WAREHOUSEING

McGraw-Hill

Estados Unidos 1998

W.H. Inmon

Building the Data Warehouse

John Wiley & Sons Editions

Estados Unidos 1996

Beatriz Jiménez S.

Almacenes de datos

Club de Investigación Tecnológica

San José, Costa Rica 1998

Bibliografía en Internet

<http://dhrinfo.hr.state.or.us/intranet/tands/Dwpap/>

http://www.cait.wustl.edu/papers/prism/vol1_no1/

<http://www.map.es/csi/silice/>

<http://www.sysameri.com/osc/wp3sp.htm>

<http://www.arcride.edu.ar/appei/>

<http://www.wnet.es/>

<http://www.fciencias.unam.mx/revista/soluciones/30s/No34/index.html>

<http://www.dwinfocenter.org/>