

**ESCUELA SUPERIOR POLITECNICA
DEL LITORAL**
ESCUELA DE POSTGRADO EN ADMINISTRACION
DE EMPRESAS

TESIS DE GRADO

**MAESTRIA EN
GESTION
ESTRATEGICA
CON EL DESARROLLO
DE
LA INFORMATICA**

DATA WARE HOUSE
CONCEPTOS Y METODOLOGIA

ELABORADO POR:

EDUARDO CALDERON V.

FERNANDO CHIANG E.

GONZALO FRIAS S.

ORLANDO RODRIGUEZ A.

CARLOS VILLACIS G.

INDICE

1. INTRODUCCION	3
2. CONCEPTOS DE DATA WAREHOUSE	4
2.1. ¿QUÉ ES DATA WAREHOUSE?	4
2.2. SISTEMAS OPERACIONALES VS SISTEMAS INFORMACIONALES	5
2.3. ANALISIS MULTIDIMENSIONAL	6
2.4. APLICACIONES	8
3. MARCO DE DESARROLLO DEL DATA WAREHOUSE	10
3.1. BASE DE DATOS OPERACIONALES Y BASE DE DATOS EXTERNAS	11
3.2. ACCESO DE INFORMACION	11
3.3. ACCESO A DATOS	12
3.4. DIRECTORIO DE DATOS	13
3.5. ADMINISTRACION DE PROCESOS	14
3.6. APLICACIONES DE MENSAJERIA	14
3.7. DATA WAREHOUSE FISICO	15
3.8. ESCENARIO DE TRANSFORMACION DE DATOS	15
4. OPCIONES DE ARQUITECTURAS PARA EL DESARROLLO DE DATA WAREHOUSE	17
4.1. ALCANCE DEL DATA WAREHOUSE	17
4.2. REDUNDANCIA DE DATOS	17
4.2.1. DATA WAREHOUSE VIRTUAL O PUNTO A PUNTO	17
4.2.2. DATA WAREHOUSE CENTRALIZADOS	18
4.2.3. DATA WAREHOUSE DISTRIBUIDOS	19
4.3. TIPO DE USUARIO FINAL	20
5. DESARROLLANDO UN DATA WAREHOUSE	21
5.1. DESARROLLANDO UNA ESTRATEGIA PARA DATA WAREHOUSE	21
5.2. EVOLUCION DE LA ARQUITECTURA DEL DATA WAREHOUSE	22
5.3. DISEÑO DEL DATA WAREHOUSE	22
5.4. ERRORES MAS COMUNES	22
6. IMPLEMENTACIÓN DEL DATA WAREHOUSE	27
6.1. SELECCION DE HARDWARE	27
6.2. SELECCION DE SOFTWARE	28
6.2.1. BASES DE DATOS	28
6.2.2. HERRAMIENTAS DE EXTRACCION	31

6.2.3. HERRAMIENTAS DE ANALISIS PARA USUARIO FINAL	32
7. ADMINISTRACION DEL DATA WAREHOUSE	34
7.1. ANÁLISIS DEL NEGOCIO	34
7.2. ADMINISTRACION DE USUARIOS	35
7.3. ADMINISTRACION DEL METADATA	37
7.4. ADMINISTRACIÓN Y SELECCIÓN DE PERSONAL	41
8. CONCLUSIONES	46
9. REFERENCIAS BIBLIOGRAFICAS	47



1. INTRODUCCION

Al finalizar los años ochenta, las empresas habían solucionado sus problemas de automatización de transacciones, esto les permitió ser mas ágiles en la recolección de información y a su vez ampliar la cantidad de información que se registraba, sin embargo a pesar de contar con un gran volumen de datos, no disponían de información oportuna para la toma de decisiones.

Un factor crítico, para sobrevivir en los noventa, es la capacidad de analizar, planificar y reaccionar rápidamente a los continuos cambios en las condiciones de los negocios, para logra esto, los ejecutivos de las empresas necesitan mas y mejor información.

La tecnología de información, ha hecho posible revolucionar la manera en que las empresas operan a través del planeta, sin embargo la triste realidad, es que a pesar de tener sofisticado equipo computacional conectado en red que abarcan todo el globo terráqueo, los ejecutivo y tomadores de decisión no disponen de información crítica que existe en la organización.

Diariamente las organizaciones crean millones de datos sobre los aspectos del negocio, datos sobre sus clientes, productos , operaciones y gente, pero en su mayor parte, estos datos están bloqueados en una maraña de sistemas de información que hace imposible su acceso, a este fenómeno lo podríamos denominar “datos enclaustrados”.

Los expertos estiman que solo una pequeña parte de los datos, que son capturados, procesados y guardados en las empresas, actualmente está disponible para los ejecutivos y tomadores de decisiones, mientras la tecnología para la manipulación y presentación de los datos se ha desarrollado aceleradamente en los últimos diez años, solo en los últimos años los desarrolladores involucrados en el desarrollo de estrategias usando tecnología de información, se han dado cuenta de la importancia de crear información bajo un esquema totalmente distinto al tradicional, a este nuevo enfoque se lo ha denominado DATA WAREHOUSE.

Este concepto de data warehouse, es un proceso bastante complejo, cada implementación tiene sus objetivos, los cuales conllevan a crear una estructura enfocada a satisfacerlos, por lo que cada proyecto tiene sus características particulares. Este trabajo está enfocado a presentar una metodología que guíe a quien inicie un proyecto de esta naturaleza, hemos tratado de abarcar todo el proceso de implementación desde el diseño de la arquitectura de hardware, selección de software, análisis de las necesidades del cliente, construcción de información y por último la administración del data warehouse, consideramos estas áreas como las de mayor importancia en el proyecto.

2. CONCEPTOS DE DATA WAREHOUSE

2.1. ¿Qué es Data Warehouse?

De acuerdo a W. H. Inmon, el padre del concepto de data warehouse, considera que "Un proyecto de Data Warehouse es la recolección de datos no volátiles, integrados, históricos y de orientación subjetiva con el fin de soportar el proceso de toma de decisiones"

Recientemente, un conjunto de nuevos conceptos y herramientas, se ha convertido en una nueva tecnología que ha permitido atacar el problema de proveer de información a las empresas, creada a partir de sus datos operacionales y datos externos, con el fin de poder sobrevivir y prosperar en un mundo cada vez mas competitivo.

El término que ha caracterizado esta nueva tecnología es DATA WAREHOUSING, este concepto ha crecido a partir de los intentos de los investigadores para proveer a sus empresas de información que les permita ser mas flexibles, efectivas y eficientes, pasando a ser el activo de mayor valor. Esta ha sido la base, en muchos casos, que ha permitido crear ventajas competitivas.

Existen muchos conceptos sobre lo que es data warehouse, para unos es una tienda de datos consistentes separada de los sistemas transaccionales, para satisfacer necesidades de reportes y acceso de datos, para otros es simplemente una arquitectura, nosotros consideramos que data warehouse es una arquitectura creada para soportar un proceso continuo de recabación de datos de fuentes heterogéneas, incluyendo datos históricos y datos sindicados para soportar la continua necesidad de requerimientos de información estructurada, reportes analíticos e información para toma de decisiones.

Data warehousing es un concepto que se basa en la integración de diferentes tecnologías, metodologías y experiencias en los últimos veinte años, estas experiencias han permitido a la industria de la tecnología de información, identificar cuales son los problemas claves que deben ser solucionados.

Existen muchas aplicaciones de data warehouse a nivel mundial, pero estimamos que el concepto es extremadamente amplio, todavía se deben optimizar muchos factores para considerar que el concepto es explotado en todo su potencial, entre otros, los siguientes factores todavía se deben desarrollar:

- Integración de datos y metadata a partir de varias fuentes.
- Calidad de datos: Limpios y refinados.
- Sincronización de las fuentes de datos que nutren el data warehouse, para asegurar un proceso continuo de refrescamiento de información, conforme se generen nuevos datos en las fuentes.
- Administración del Metadata.

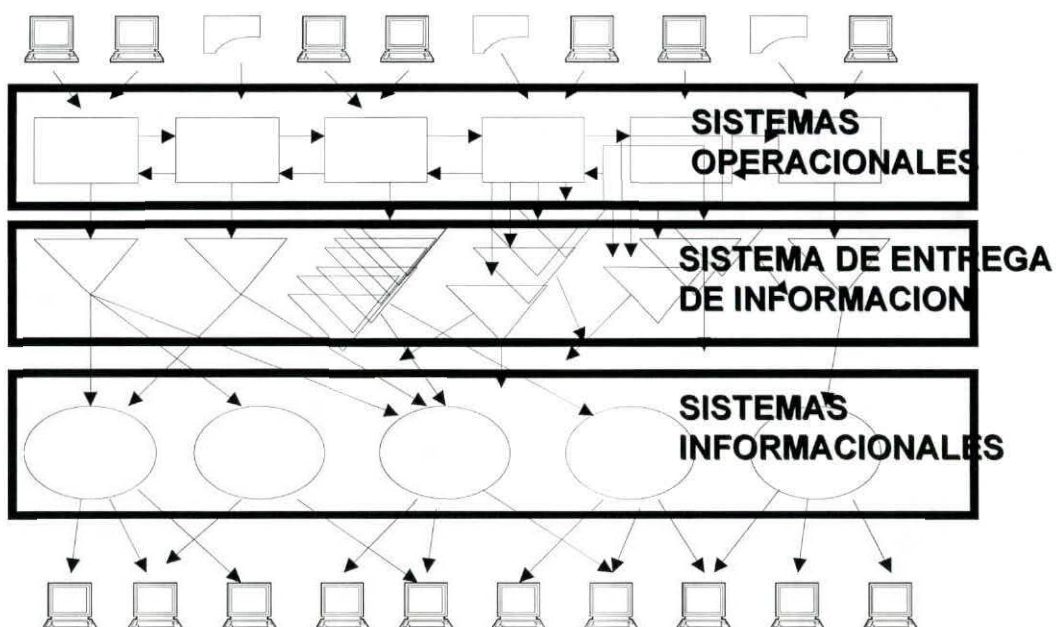
Como podemos apreciar, data warehouse es un concepto amplio, que debe ser analizado y aplicado en función de los objetivos de cada proyecto.

2.2. SISTEMAS OPERACIONALES VS SISTEMAS INFORMACIONALES

Uno de los conceptos mas importantes que ha surgido del desarrollo de data warehouse, es el reconocimiento que fundamentalmente existen dos tipos de sistemas de información en las organizaciones: Sistemas operacionales y sistemas informacionales.

Sistemas operacionales, son los sistemas que ayudan a procesar las transacciones diarias de la empresa, como es el caso de los sistemas de inventario, contabilidad, cuentas por cobrar, facturación, etc. Debido a su importancia para la organización, estos sistemas siempre son los primeros que se implementan en una empresa, al pasar los años, estos sistemas han sido extendidos y rediseñados, a tal punto que han integrado totalmente la operación de la empresa, tanta es su importancia que muchas empresas en la actualidad no podrían operar sin estos sistemas y los datos que contiene.

Por otro lado, existen otras funciones que se realizan en la empresa como la planificación, presupuesto y administración de la misma, esta funciones también son críticas para la sobrevivencia de la organización, especialmente en el entorno tan cambiante que existe actualmente, funciones como planificación de métodos de mercadeo y análisis financiero, requieren de sistemas de información para poder realizarlas, sin embargo estas son diferentes a las operacionales, y los tipos de sistemas e información requerida son también diferentes. Las funciones basadas en conocimiento son sistemas informacionales.



Los sistemas informacionales dan apoyo al análisis de información y toma de decisiones, inclusive decisiones trascendentales sobre como la empresa debe operar, ahora y en el futuro. Los sistemas informacionales no solo tienen un enfoque diferente a los operacionales, tienen un alcance distinto, los operacionales normalmente se enfocan a un área específica, los informacionales relacionan diferentes áreas, necesitando una gran cantidad de datos operacionales.

En los últimos años, data warehousing se ha desarrollado aceleradamente, pasando de ser un conjunto de ideas en una arquitectura para creación y entrega de información.

2.3. ANALISIS MULTIDIMENSIONAL

La manera natural en que los usuarios consultan información, es usando su forma de expresión normal, por ejemplo necesitan saber el total de ventas por producto en la zona norte del país, esta forma de expresarse dista mucho del lenguaje técnico que se utiliza para obtener esta información.

Desde una perspectiva de análisis, los usuarios desean extraer datos de una forma intuitiva, con una mínima inversión en tiempo y con el menor grado de frustración. Después de obtener los datos, ellos analizan, sintetizan y consolidan estos en información, la cual usarán para la toma de decisiones.

Existen varias formas de extraer información de valor del data warehouse para su posterior análisis, en la siguiente gráfica analizamos este proceso



En el proceso de soporte a las decisiones, el proceso analítico predominantemente se suscribe a la verificación del modo de usarlo, en el modo de verificación, el usuario crea una hipótesis, en otras palabras, una pregunta de negocio y a partir de esta accesa datos para verificar o confirmar la hipótesis. El análisis de datos es usualmente iterativo, una pregunta antecede a otra, la que a su vez genera otra, hasta que se deduce un grupo de alternativas y recomendaciones. Durante la actividad de análisis iterativo, usuarios experimentados a veces descubren relaciones insospechadas entre parámetros de negocios, por ejemplo, la relación de ventas con el sexo y la edad de los clientes, esto no implica que un proceso de verificación siempre llevará a un descubrimiento, este se genera a partir de un procesamiento de observación del usuario.

Datos del negocio, son por naturaleza multidimensionales, están interrelacionados y usualmente son jerárquicos, por ejemplo datos de ventas, datos de inventario y el presupuesto son interrelacionados e interdependientes. En la práctica, el presupuestar la venta de nuevos productos, análisis del patrón de compras histórico, adopción de nuevos productos, preferencias regionales, y otros factores del negocio, requiere análisis.

Tanto como para el control de la eficiencia operacional como para el de planificación, se deben analizar muchos datos interrelacionados del negocio, esta necesidad se la realiza a través del procesamiento analítico, en el cual el enfoque es el de análisis de datos, específicamente el de análisis multidimensional.

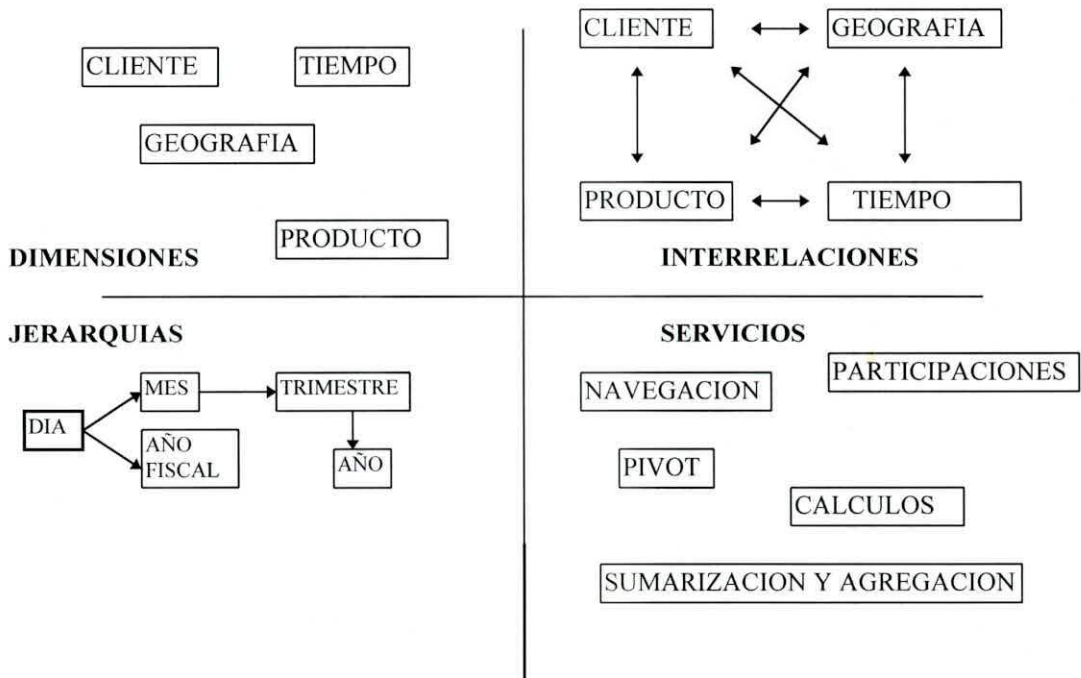
En análisis multidimensional, los datos son representados como dimensiones tal como productos, territorios y clientes. Las dimensiones están usualmente relacionadas en jerarquías, por ejemplo ciudad, estado, región, país y continente, o en estado territorio y región. Dimensiones de tiempo son dimensiones estándar con su propia jerarquía, tal como día, semana, mes, trimestre y año, o día y acumulado del año.

Estos esquemas de análisis complejos, procesamiento analítico o análisis multidimensional, presentan los datos en un formato muy amigable, un usuario puede acceder la utilidad por departamento y tienda durante los últimos cuatro trimestres para un grupo de productos. Los resultados pueden ser pivoteados o rotados para cambiar los ejes y las perspectivas, adicionalmente los usuarios pueden navegar dentro de la información a niveles de mayor sumariazación o bajar a niveles de mayor detalle, así como también a través de las dimensiones para analizar otras perspectivas. Otras facilidades del análisis multidimensional le permiten ranquear los productos tope o los últimos en función de los parámetros definidos por el usuario, por ejemplo dame los cinco mejores vendedores.

El procesamiento analítico es usado para comprender que es lo que pasa en el negocio, permitiendo análisis de sensibilidad.

El alcance del procesamiento informacional, es usualmente en dos o tres dimensiones, se enfoca a el análisis de datos históricos para entender el pasado, el procesamiento analítico puede ser usado para análisis históricos complejos con manipulación extensiva de análisis de datos, así como para planificación y presupuestos partiendo de datos históricos para predecir el futuro. El escoger erróneamente el enfoque de procesamiento informacional o el de procesamiento analítico, puede llevar a frustrarse al usuario, causando un negativo impacto en la productividad y como consecuencia la pérdida de una ventaja competitiva estratégica.

El siguiente gráfico detalla los diferentes elementos del análisis multidimensional.



2.4. APLICACIONES

Un proyecto de data warehouse, puede ser creado para un fin específico o para cumplir con varios fines, incluso debe estar estructurado para satisfacer nuevas necesidades, esto implica una estrategia de uso de recursos y de administración dependiente del objetivo del proyecto, a continuación mencionaremos algunos ejemplos de data warehouse:

- Una cadena de venta de zapatos, necesitaba establecer cuales eran las costumbres de compra de sus clientes, su cadena de distribución, consistía en varias tiendas pequeñas que reportaban a una oficina central el movimiento diario, crearon un data warehouse que analizaba históricamente la conducta de cada cliente, estableciendo de esta manera un profundo conocimiento de los hábitos de compra del mismo, esta información se la enviaron a las tiendas y estas pudieron mercadear al cliente de una forma mas agresiva, como ofreciéndoles los modelos y colores que el cliente habitualmente compraba.
- Una empresa de distribución de gaseosas, necesitaba determinar cuales eran los niveles óptimos de producción de las siete marcas de refrescos que fabricaba, para lograr este objetivo, creó un data warehouse para analizar los hábitos de compra de sus clientes, de esta manera estableció los volúmenes de compra y sus frecuencias por marca y zonas, con lo que bajó sus niveles de inventario y mejoró la eficiencia del mix de productos que transportaba cada camión.
- Un banco cambió su estructura jerárquica a horizontal, creando gerencias de productos que le permitía ser mas ágil para adaptarse a los continuos cambios que

exigía el entorno, sin embargo la estructura de sistemas no era lo suficientemente ágil para satisfacer las necesidades de información de estas gerencias, por lo que conformó una unidad para el desarrollo del data warehouse del banco con el fin de que estos requerimientos sean atendidos de una forma rápida y oportuna.

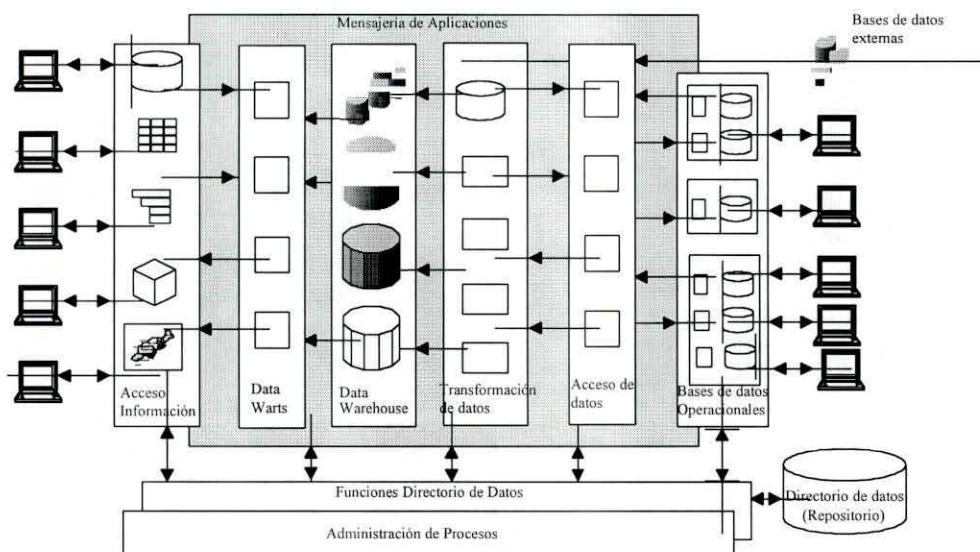
Cada uno de estos ejemplos tiene en común, la necesidad de crear información a partir de datos transaccionales, para reducir la incertidumbre en el proceso de toma de decisiones, sin embargo las necesidades de las empresas son muy variables y específicas, por lo que cada aplicación del concepto de data warehouse, con toda seguridad será muy diferente.

3. MARCO DE DESARROLLO DEL DATA WAREHOUSE

Una de las razones por las que el data warehouse ha tomado tanto tiempo en desarrollarse, ha sido el lento proceso del desarrollo de tecnología para poder lograr los objetivos del data warehouse, si bien es cierto se ha avanzado mucho, todavía falta el desarrollo de nuevas tecnologías que permitan cumplir totalmente con este concepto, como tal, data warehouse puede ser representado como un marco de desarrollo que abarca toda la organización, para crear y administrar información. Para poder entender como todos los componentes involucrados en la estrategia de data warehousing están relacionados, es esencial tener una arquitectura de data warehouse.

La arquitectura del data warehouse, es una manera de representar toda la estructura de datos, comunicaciones, procesos y presentación que existe para los usuarios finales del sistema dentro de la empresa. Esta arquitectura esta compuesta de varias partes interconectadas:

- Nivel de base de datos operacional y base de datos externas.
- Nivel de acceso de información.
- Nivel de acceso de datos.
- Nivel de directorio de datos (metadata).
- Nivel de administración de procesos.
- Nivel de aplicaciones de mensajería.
- Nivel del data warehouse.
- Nivel del escenario de transformación de datos.



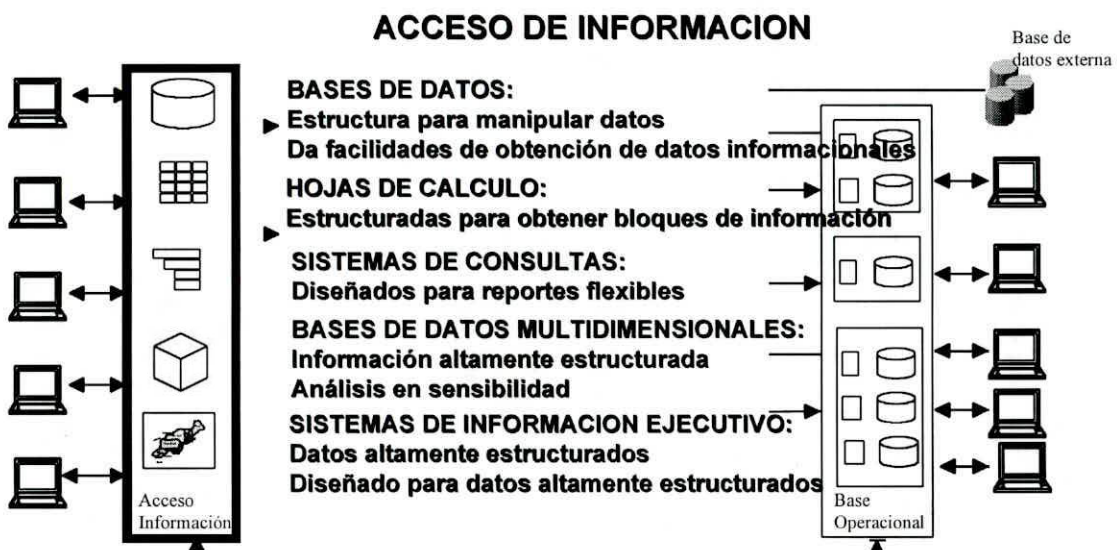
3.1. BASE DE DATOS OPERACIONALES Y BASE DE DATOS EXTERNAS

Sistemas operacionales, procesan datos para soportar necesidades operacionales críticas, para realizar esto, base de datos operacionales, han sido históricamente creadas para proveer una eficiente estructura de proceso para un relativo pequeño número de transacciones de negocio bien definidas, sin embargo, debido al enfoque limitado de los sistemas operacionales, las bases de datos diseñadas para soportar sistemas operacionales, son difíciles de usar para acceder datos con propósitos administrativos o informacionales. Esta dificultad para acceder datos operacionales, se torna más crítica si tomamos en cuenta de que usualmente tienen entre diez y quince años de vida, esto implica que la tecnología usada con toda seguridad es obsoleta.

El objetivo de data warehousing, es liberar la información que esta enclaustrada en las bases de datos operacionales y mezclarla con información externa, tendencia cada vez más creciente por parte de las empresas, que se han dado cuenta de la importancia de analizar información externa, esta puede ser datos demográficos, econométricos, tendencias de compras y de la competencia, etc.

3.2. ACCESO DE INFORMACION

El nivel de acceso de información de la arquitectura del data warehouse, es el nivel con que el usuario final trata directamente, en particular, representa las herramientas que normalmente usa día a día, como por ejemplo Excell, Lotus, Focus, Access, SAS, etc, este nivel también incluye el hardware y software involucrado en presentar e imprimir reportes, hojas de cálculo y gráficos para análisis y presentaciones, durante los últimos veinte años, este nivel se ha desarrollado enormemente, especialmente cuando los usuarios finales comenzaron a usar PC's y redes.



Actualmente, existe una gran variedad de herramientas sofisticadas para manipular, analizar y presentar información, sin embargo, hay varios problemas en el proceso de obtención de los datos contenidos en sistemas operacionales, al momento de ser requerido por estas herramientas, la complejidad que se genera debido a lo heterogéneo de los ambientes que contienen los datos operacionales, es usualmente uno de los mas grandes problemas a resolver, esto puede complicarse mas si estamos hablando de sistemas distribuidos ubicados en lugares físicos distantes. La tecnología en la actualidad ha solucionado parte de este problema, pero todavía existen ciertas complicaciones para obtener esta información, una de las claves es el encontrar un lenguaje común de datos que pueda ser usado a través de toda la empresa.

3.3. ACCESO A DATOS

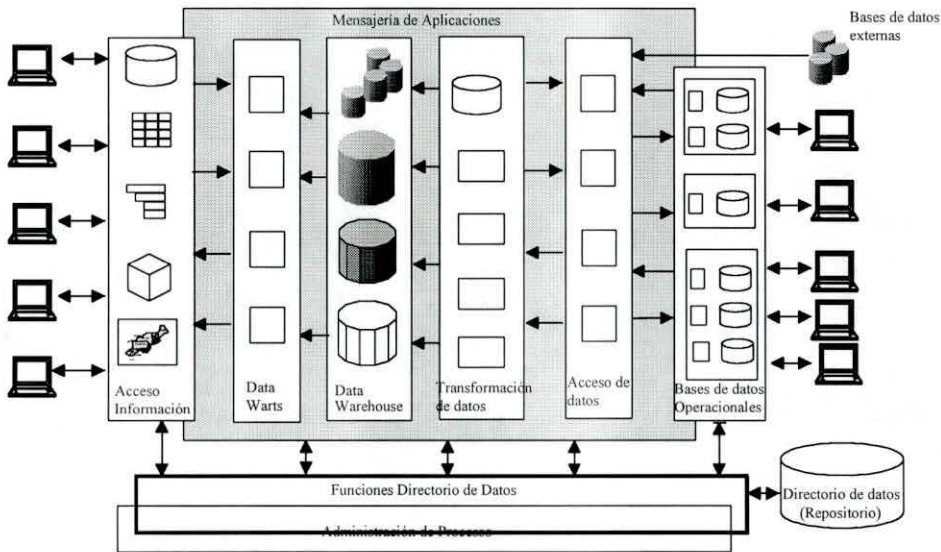
El nivel de acceso de datos en la arquitectura de data warehouse, es el que permite al nivel de acceso de información conversar con el nivel operacional, en el mundo de las redes, en la actualidad el lenguaje común que se ha desarrollado es el SQL (Structured query language), lenguaje que originalmente fue desarrollado por IBM como un lenguaje de consulta, sin embargo durante los últimos veinte años se a convertido en el estándar de facto para intercambio de datos.

Una de las herramientas tecnológicas desarrollada en los últimos años, que ha impactado tremendamente en el desarrollo del concepto de data warehouse, es una serie de "filtros", que hacen posible a través de estatutos SQL acceder prácticamente todas las bases de datos y sistemas de archivos de datos, relacionales y no relacionales. Estos filtros han hecho posible el desarrollo de herramientas de acceso de información eficientes al momento de acceder datos guardados en sistemas de administración de base de datos de veinte años de antigüedad.

El nivel de acceso de datos, no solo cubre diferentes base de datos y sistemas de archivo bajo el mismo hardware, también cubre diversos constructores y protocolos de red, logrando de esta manera un acceso de datos independiente del hardware, software y protocolos de red con que se conectan las computadoras que conforman el WAN (wide area network) de la organización, a esto lo denominamos "acceso de datos universal", cuyas característica principal, es la de permitir al usuario final el acceso a los datos de la organización independientemente de donde se encuentren.

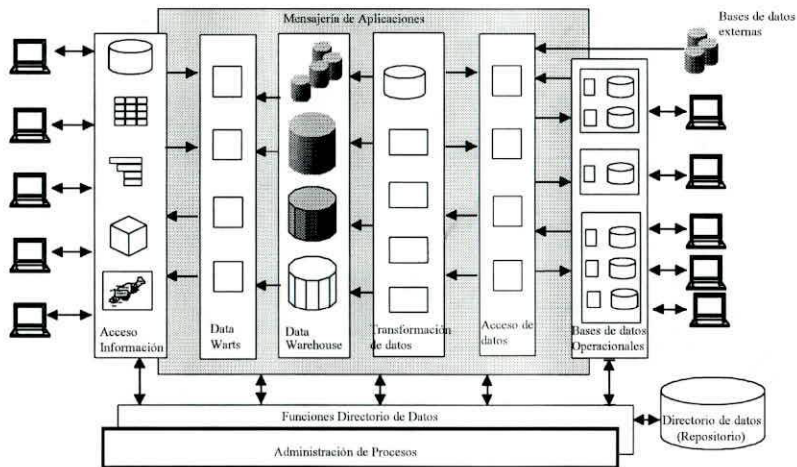
3.4. DIRECTORIO DE DATOS

Para lograr el objetivo de acceso de datos universal, es absolutamente necesario el mantener un directorio de datos o repositorio de metadata (datos acerca de los datos). Su función es la de registrar las características de la información que se encuentra almacenada en el data warehouse, es importante también registrar información sobre el punto de vista de los usuarios finales, datos sobre las base de datos operacionales, perfiles de usuarios, herramientas de usuario final asignadas, etc. Lo ideal sería que los usuarios finales tuviesen la información necesaria para poder acceder información sin tener que saber donde reside la información o la forma en que fue almacenada.



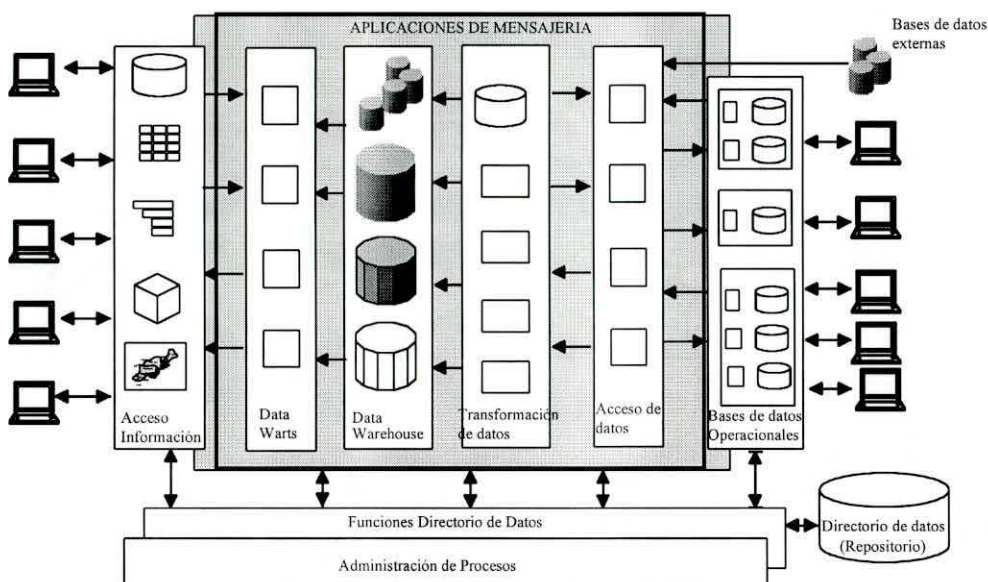
3.5. ADMINISTRACION DE PROCESOS

El nivel de administración de procesos está involucrado en la programación de las diversas tareas que deben ser cumplidas, para construir y mantener la información del data warehouse y el metadata. Este nivel puede ser conceptualizado como el controlador de alto nivel, de todos los procesos o procedimientos que deben realizarse para mantener el data warehouse actualizado.



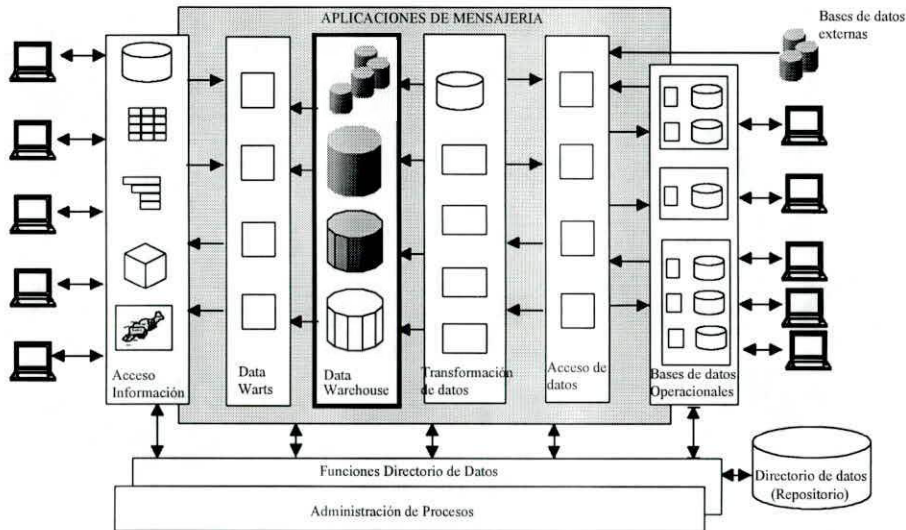
3.6. APLICACIONES DE MENSAJERIA

El nivel de aplicaciones de mensajería, tiene que ver con el transporte de información a través de la red computacional de la empresa, también se lo denomina "middleware", este nivel, puede involucrar más que protocolos de red, puede realizar operaciones de recolección de transacciones o mensajes y enviarlos a ciertas localidades en un momento determinado.



3.7. DATA WAREHOUSE FISICO

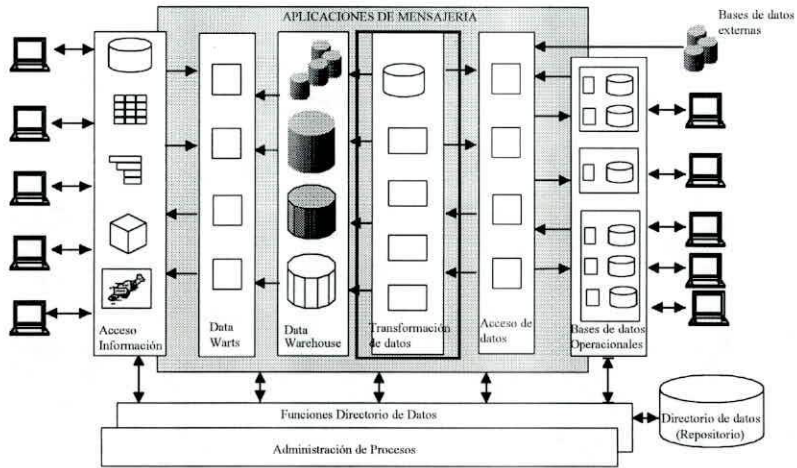
El nivel físico del data warehouse, es donde se encuentra los datos usados para usos informacionales, en algunos casos, se puede pensar sobre el data warehouse como una vista lógica o virtual de los datos, en algunos casos el data warehouse no esta involucrado en el almacenamiento de información.



En este nivel, copias de datos operacionales internos y externos son almacenados de tal manera que es muy fácil de acceder y extremadamente flexible de usar. Usualmente se los almacena en plataformas cliente servidor, pero también han sido almacenados en main frames.

3.8. ESCENARIO DE TRANSFORMACION DE DATOS

El componente final de la arquitectura de data warehouse, es el escenario de transformación de datos, también se lo conoce como administrador de copia o replicación, pero en realidad incluye todos los procesos necesarios para seleccionar, editar, sumarizar, combinar y cargar los datos del data warehouse y el metadata a partir de las bases de datos operacionales y externas.



Este escenario involucra programación compleja, sin embargo actualmente se dispone de varias herramientas para simplificar este proceso. Este nivel también incluye programas de análisis de calidad de datos, así como también filtros que identifiquen patrones y estructuras de datos dentro de los datos operacionales.

4. OPCIONES DE ARQUITECTURAS PARA EL DESARROLLO DE DATA WAREHOUSE

Existen muchas maneras de desarrollar un proyecto de data warehouse, en realidad podría haber una por cada empresa, sin embargo existen diferentes dimensiones que deben ser consideradas:

- Alcance del data warehouse.
- Redundancia de datos.
- Tipo de usuario final

4.1. ALCANCE DEL DATA WAREHOUSE

El alcance del data warehouse, puede ser tan amplio como la cobertura de todos los datos informacionales de la organización desde su inicio, o puede ser tan pequeño como una data warehouse personal para un ejecutivo con información del año en curso, a pesar de todo, los dos son conceptos de data warehouse. En la práctica, mientras más amplia es la cobertura, el data warehouse tiene un valor mayor a la empresa, y a su vez es más costoso y requiere más tiempo de desarrollo para su creación y mantención, como consecuencia, muchas empresas inician sus proyectos con data warehouse departamentales o divisionales, y a partir de estos se expanden conforme sus usuarios lo requieran.

4.2. REDUNDANCIA DE DATOS

Existen esencialmente tres niveles de redundancia de datos sobre los que la empresa debe pensar cuando considera sus opciones de data warehouse:

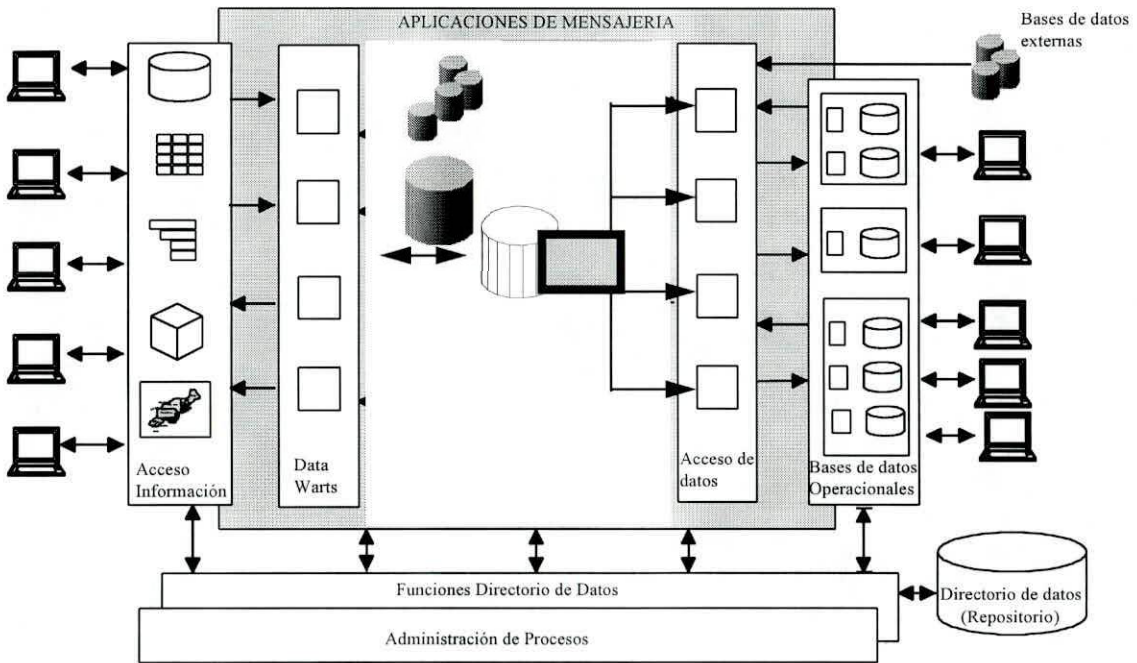
- Data warehouse "virtual" o "punto a punto".
- Data warehouse centralizados.
- Data warehouse distribuidos.

No existe la mejor opción, cada una encaja de acuerdo a las necesidades de la empresa, y es posible que una estrategia de data warehouse incluya las tres opciones.

4.2.1. DATA WAREHOUSE VIRTUAL O PUNTO A PUNTO

Una estrategia de data warehouse virtual o punto a punto, significa que los usuarios finales acceden a las bases de datos operacionales directamente usando herramientas configuradas para satisfacer sus necesidades de información. Este esquema, provee una gran flexibilidad así como el mínimo nivel de redundancia en los datos que deben ser cargados y mantenidos, también es verdad que esta opción, podría en un

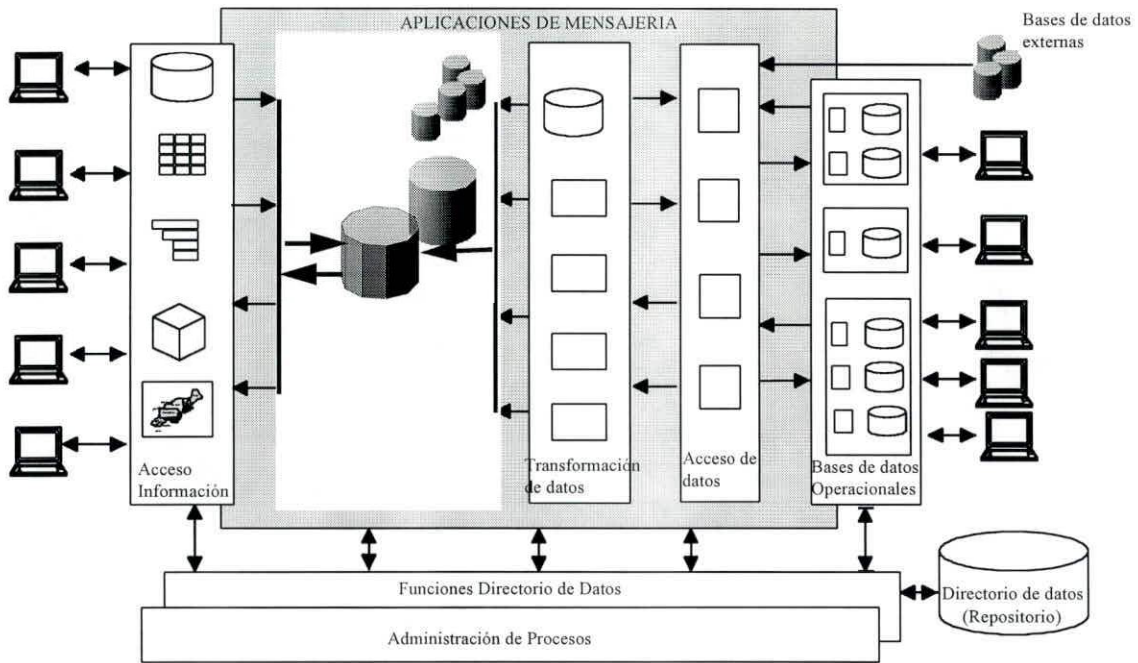
momento dado, consumir extremadamente recursos del computador a través de una consulta compleja.



Un data warehouse virtual, usualmente es la estrategia inicial en empresas donde existe una gran cantidad de usuarios cuyas necesidades no están definidas, y cuya frecuencia de requerimientos de información es baja, este esquema provee un punto de partida, para que las organizaciones aprendan lo que los usuarios finales realmente quieren.

4.2.2. DATA WAREHOUSE CENTRALIZADOS

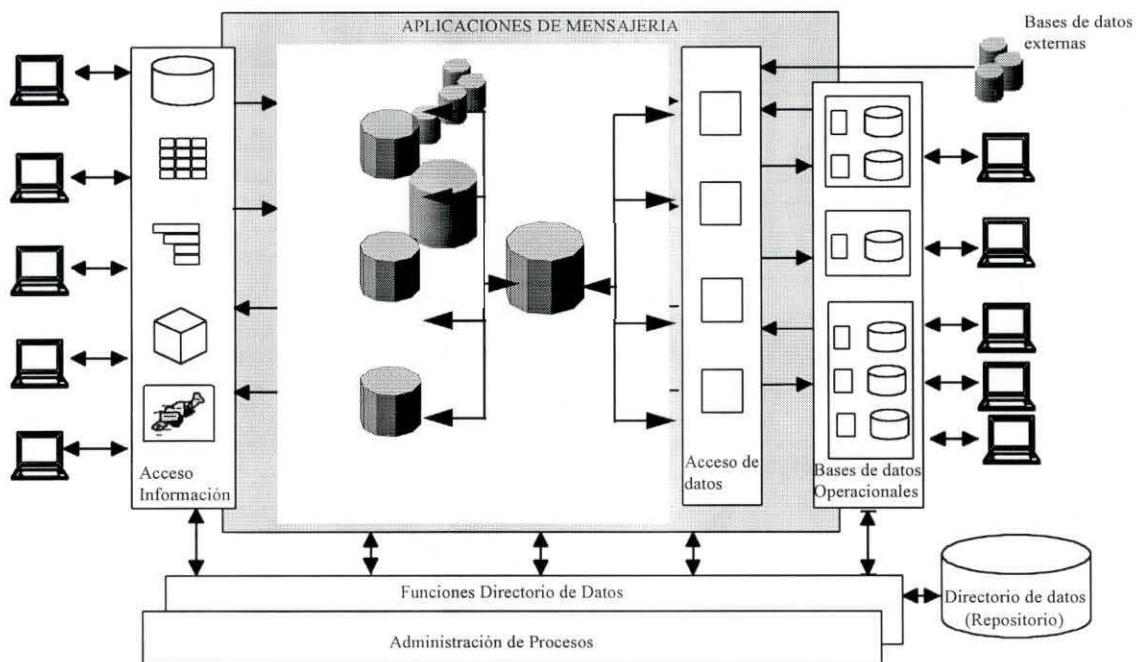
Data warehouse centralizados, es el concepto adoptado por la mayoría de las personas que se inician en este proceso, este esquema de centralización tiene una base de datos donde reside la información de un área funcional, departamento, división o empresa, usualmente se la implementa donde existe necesidades comunes de información y a su vez, los usuarios están conectados a una computadora central o red.



Los data warehouse centralizados, son reales, los datos almacenados son accesible a partir de un solo lugar, y deben ser cargados y mantenidos en períodos de tiempo definidos, normalmente son construidos sobre software avanzado de base de datos relacionales, o sobre base de datos multidimensionales.

4.2.3. DATA WAREHOUSE DISTRIBUIDOS

Data warehouse distribuidos, son aquellos en los cuales ciertos componentes están distribuidos a través de varias bases de datos físicas. Las empresas, constantemente presionan a niveles mas bajos a que tomen decisiones, por ende las necesidades de información de estas personas debe satisfacerse a través de redes locales o computadoras personales.



Data warehouse distribuido, es el esquema que involucra el mayor nivel de datos redundantes, y como consecuencia su carga y actualización son mas complejas.

4.3. TIPO DE USUARIO FINAL

De la misma manera en que existen diversas formas de organizar el data warehouse, es importante analizar, el que existe un amplio rango de usuarios finales, para su clasificación los encasillaremos en tres categorías:

- Ejecutivos y administradores.
- Usuarios intensivos (analistas de negocio y financieros, ingenieros, etc).
- Usuarios de soporte (administrativos, etc)

Cada una de estas diferentes categorías de usuario tiene sus propios requerimientos de datos, acceso, flexibilidad y facilidad de uso.

5. DESARROLLANDO UN DATA WAREHOUSE

Desarrollar un buen data warehouse, no es diferente al desarrollo de otro proyecto de tecnología de información, este requiere una planificación cuidadosa, definición de requerimientos, diseño, construcción de prototipos e implementación. El primer y más importante elemento es el proceso de planeación, el cual determina que tipo de estrategia de data warehouse debe seguir en sus inicios.

5.1. DESARROLLANDO UNA ESTRATEGIA PARA DATA WAREHOUSE

Antes de desarrollar un data warehouse, es crítico el desarrollar una estrategia balanceada de data warehousing, que cumpla con las necesidades de la empresa y sus usuarios. Se debe determinar quienes son los usuarios, alcance, que tipo de data warehouse se debe desarrollar, etc.

Se puede definir varias estrategias para iniciar un proyecto de data warehouse, una de ellas es establecer un ambiente de data warehouse virtual, el cual se podría crear siguiendo estos pasos:

1. Instalar los siguientes niveles: Acceso de datos, directorio de datos y facilidades de administración de procesos.
2. Entrenar los usuarios finales.
3. Monitorear el uso de las facilidades de data warehouse.
4. Basado en el uso actual, crear un data warehouse físico que soporte los requerimientos de alta frecuencia.

Una segunda estrategia, podría ser construir una copia de los datos operacionales en otro computador y habilitar el data warehouse utilizando una serie de herramientas para acceso de información. Esta estrategia tiene la ventaja de ser simple y rápida, desgraciadamente si los datos son de mala calidad y/o el acceso a los datos es extremadamente complejo, este esquema podría crear serios problemas.

Una buena estrategia para diseño de data warehouse, es seleccionar una población representativa de las diferentes áreas de la empresa y analizar sus necesidades de información, basado en este análisis, desarrollar un prototipo de data warehouse, a partir del cual los usuarios experimentarán y afinarán sus requerimientos. Cuando se hallan convertidos en usuarios maduros, se podrá construir un data warehouse definitivo y abrirlo a toda la empresa. Si se requiere, se pueden habilitar herramientas de acceso de información, para permitir a usuarios finales el acceso a los datos requeridos usando sus herramientas favoritas, o utilizar herramientas de acceso multidimensional usando la información básica del data warehouse como base.

No existe un esquema de construcción de data warehouse que satisfaga las necesidades de todas las empresas, las necesidades de cada una son diferentes de acuerdo a su contexto, por último conforme se van desarrollando nuevos proyectos, se crean nuevas técnicas de desarrollo por lo que este concepto de data warehouse continuamente está evolucionando.

5.2. EVOLUCION DE LA ARQUITECTURA DEL DATA WAREHOUSE

Es muy complicado juntar todos los niveles de la arquitectura del data warehouse en sus inicios, esto se puede lograr estableciendo un plan que permita ir construyendo esta arquitectura en etapas, este desarrollo junto a un claro conocimiento de los objetivos del proyecto, permitirá determinar si la construcción del data warehouse está encuadrado en un marco razonable.

Una de las claves del data warehouse, es su flexibilidad, es crítico el tener en cuenta que un proyecto exitoso de esta naturaleza, atraerá una gran cantidad de usuarios finales.

5.3. DISEÑO DEL DATA WAREHOUSE

El diseño de data warehouse es muy diferente del diseño de los sistemas operacionales tradicionales, comúnmente los usuarios de data warehouse no conocen sus necesidad ni lo que quieren, contrario a los usuarios operacionales que generalmente conocen profundamente su actividad, por otro lado el diseño de data warehouse involucra pensar en términos más amplios y más difíciles de definir, conceptos de negocio, etc, en los sistemas operacionales esto es mucho mas sencillo ya que están enfocados a necesidades puntuales, finalmente la estrategia ideal de diseño para data warehouse es de afuera hacia adentro, contrario a la de sistemas operacionales que es de arriba hacia abajo.

En realidad todas estas diferencias no son tan importantes a la hora del diseño, como el hecho de poder orientar a los usuarios a que encuentren cuales son sus verdaderas necesidades, el diseñador debe utilizar varios métodos o trucos para poder definir por cada usuario que es lo que realmente quiere, y esto será un factor decisivo para el éxito del proyecto.

5.4. ERRORES MAS COMUNES

Data warehouse como concepto es relativamente nuevo y en constante evolución, hemos visto que su arquitectura depende específicamente de las características de la empresa, sus recursos y sus usuarios, ante la amplitud en cuanto al criterio que se aplica a este concepto, el equipo responsable del proyecto puede cometer errores que comprometen el éxito del mismo, a continuación presentamos los errores mas comunes:

- **Auspicio adecuado de los altos ejecutivos:** Todo proyecto de data warehouse, debe tener apoyo por parte de ejecutivos de mayor nivel que el responsable del proyecto, se recomienda que por lo menos sean dos ejecutivos quienes impulsen el proyecto, uno de ellos debe ser un alto ejecutivo con poder de decisión sobre las inversiones que se deben realizar, por otro lado cuando un proyecto de esta naturaleza falla, usualmente es por falla en la comunicación entre el auspiciante del proyecto y el responsable, asumamos que el director general de un corporativo decide iniciar un proyecto de data warehouse, al nombrar al responsable del mismo, escoge a un funcionario que le resulta muy difícil comunicarse con el director, con toda seguridad este señor va a tener dificultades durante el desarrollo, las cuales no va a poder solucionarlas por la falta de comunicación con el director general, en estos casos es importante que exista un ejecutivo intermedio como el vicepresidente de operaciones con quien el responsable pueda comunicarse, a su vez este ejecutivo tiene mayor facilidad de comunicación con el director general. A este ejecutivo normalmente se lo denomina el conductor del proyecto, y es recomendable que tenga las siguientes características:

 - Debe haberse ganado el respeto de los demás ejecutivos
 - Debe ser muy inclinado al uso de nueva tecnología
 - Con un criterio flexible, pero decidido
- **Expectativas que no se pueden cubrir y ejecutivos frustrados:** Proyectos de data warehouse tienen por lo menos dos fases: La fase de venta, en la que se trata de persuadir a los futuros usuarios, que ellos tendrán acceso a información a través de una herramienta gráfica muy sencilla de usar. La segunda fase, es la lucha por cumplir las expectativas que se crearon en la primera fase. La información almacenada en el data warehouse, no necesariamente contiene toda la información que los usuarios necesitan, la creación de esta información esta enfocada en función de objetivos específicos establecidos por los usuarios, peor aún la mayor parte de esta información es sumariada no de detalle, esto implica que si un ejecutivo a quien se le han creado expectativas sobredimensionadas, del tipo de información que puede encontrar en el data warehouse, se le contesta que la información que el requiere no está disponible, y crearla va a tomar su tiempo, con toda seguridad se sentirá frustrado, engañado y perderá el interés en el uso de la información del data warehouse.
- **Falta de política a la hora de promocionar el data warehouse:** Es muy común cometer el error de vender las bondades del data warehouse usando expresiones que puedan generar reacciones negativas, contra el proyecto y quien lo dirige, por ejemplo si se expresa que entre los beneficios del data warehouse es brindar ayuda a los ejecutivos a que tomen mejores decisiones, probablemente algunos se sentirán ofendidos porque se esta sugiriendo que actualmente no toman buenas decisiones, con toda seguridad ese usuario estará en contra del proyecto. Es importante recalcar a los usuarios finales que la intención del data warehouse, es brindar información para reducir el grado de incertidumbre en sus decisiones.
- **Crear información en el data warehouse sin un proceso exhaustivo de las necesidades de un usuarios:** En muchos proyectos de data warehouse, se comete el error de enviarles a los usuarios una lista de los elementos que conforman las base de datos operativas y externas, preguntándoles que información quisieran que se genere, con toda seguridad les van a pedir que

califiquen esa información como “esencial”, “importante”, y “sería interesante tenerla”, el resultado será una larga lista de requisitos no estructurados por parte de los usuarios, que al momento de ser creado y puesto a disposición de los usuarios, será información irrelevante frente a las verdaderas necesidades de los usuarios. El crear información de esta manera, causará el ocupar tiempo de proceso, de recurso humano, de espacio de almacenamiento en el computador y hará que la información que verdaderamente es relevante sea de mas lento acceso.

- **Crear que el desarrollo de un data warehouse es similar al diseño de una base de datos transaccional:** Data warehouse es fundamentalmente diferente del proceso de transacciones, el objetivo es acceder información agregada, sumariada, promediada, tendencias, etc. Otra diferencia es el tipo de usuario, en procesos transaccionales, un programador desarrolla una consulta que será usada muchísimas veces, en data warehouse, un usuario desarrollará una consulta que probablemente solo sea usada una vez. Las bases de datos de data warehouse normalmente son desnormalizadas, para que los usuarios al realizar sus consultas, tengan un buen tiempo de respuesta. Otra diferencia fundamental, es el contenido, sistemas transaccionales usualmente contienen datos básicos, en el data warehouse el criterio primordial de desarrollo es el de presentar la información bajo un criterio de agregados y series de tiempo, precalculados para los usuarios y listos para accederlos rápidamente, este es el criterio que ha impulsado el desarrollo de base de datos multidimensionales.
- **Escoger un administrador de data warehouse, el cual sea orientado a tecnología antes que a necesidades del usuario:** Un factor primordial para el éxito de un data warehouse, es el que se analice profundamente las necesidades del usuario, si el enfoque principal del proyecto es el uso eficiente de tecnología y se descuida el análisis del usuario, con toda seguridad ese proyecto fracasará, quien maneje el proyecto, debe ser una persona que tenga conocimientos generales del negocio y tengan buena comunicación con los usuarios, de otra manera la relación con ellos con toda seguridad será problemática y poco efectiva. El negocio de data warehouse es el de servicio de información, no almacenamiento de información, y no satisfacer las necesidades del cliente es una manera segura de salirse del negocio.
- **Enfocarse a la tradicional fuente de información interna transaccional e ignorar el valor potencial de datos externos y uso de texto, imágenes, vídeo y sonido:** Es muy importante analizar el impacto en la toma de decisiones que pueda tener la información externa, el administrador del data warehouse debe analizar junto con los usuarios cual sería el beneficio de crear información a partir de datos externos, esto debe ser analizado muy detenidamente para determinar objetivos concretos en el uso de esta información. Por otro lado el uso de multimedia para complementar la información almacenada es una de las tendencias crecientes en el desarrollo de proyectos de data warehouse.
- **Crear información con criterios no bien definidos y unificados:** El talón de aquiles del data warehouse, es el requerimiento de que exista un consenso general en las definiciones de datos, la diversidad de criterios en las definiciones de datos, normalmente son difíciles de reconciliar, esto obviamente causa problemas al momento de crear información ya que esta no será analizada bajo la misma óptica. Muchas definiciones son creadas por los usuarios bajo el criterio que mejor presente su gestión, por ejemplo si se desea presentar los niveles de colocación de dinero en un banco, probablemente pedirán que los valores colocados incluyan los

intereses, para de esta manera reflejar una mejor gestión, por otro lado también existe los diferentes puntos de vista, para el departamento de ventas, ventas significa el valor de comprado por los clientes, para el departamento de finanzas, significa ingresos menos egresos y para el departamento de distribución es tener lo necesario para realizar la entrega, en este caso existen tres definiciones distintas de “ventas”. Generalmente los ejecutivos no quieren ceder en cuanto a sus puntos de vista, solucionar este problema es probablemente una de las tareas mas importantes para el conductor del data warehouse, si no se soluciona, los usuarios no tendrán confianza en la información que están obteniendo, peor aún probablemente se van a aturdir al usar información errónea, en cuyo caso, inevitablemente odiaran al data warehouse

- **Crear en las promesas de perfomancia, capacidad y escalabilidad:** Tres empresas de distinta actividad, con arquitecturas de data warehouse totalmente diferentes, tuvieron una experiencia en común, las tres dentro de los primeros cuatro meses de haber iniciado el proyecto, cada administrador inesperadamente tuvo que comprar por lo menos un computador de igual tamaño o mas grande que el computador principal que se había definido para el proyecto, sencillamente necesitaban mas poder de proceso o almacenamiento, dos de ellos fallaron en su presupuesto por este motivo y el tercero afortunadamente había presupuestado un valor para imprevistos. Un problema común es el de capacidad en las redes. Una compañía reportó que en el diseño de su red, previó un data warehouse de imágenes, pero descubrió que la red fue rápidamente saturada, la sorpresa fue que el problema no fue la transferencia de las imágenes, sino el tráfico causado en la red por la transferencia de datos entre la aplicación del usuario final y la base de datos de índices en el servidor, las imágenes se transferían rápidamente, pero el proceso de encontrar la correcta, saturaba la red. Sobrecarga en la red, es muy común encontrar en una arquitectura cliente/servidor, especialmente en data warehouse.
- **Crear que cuando el data warehouse está implementado y operando, todos los problemas han sido solucionados:** Cada usuario feliz, generalmente va a requerir nueva información, y comentará de lo maravilloso de la nueva herramienta y obviamente quienes escucharon los comentarios requerirán mas información, pero sobre todo la quieren inmediatamente, paralelamente los problemas de perfomancia y entrega de información resulta en una fuerte presión para buscar nuevas tecnologías con que solucionarlos. Esto implica que el equipo de data warehouse debe estar preparado para operar durante un largo tiempo, es un error considerar que un proyecto de data warehouse tiene un tiempo de desarrollo y que posteriormente operará por si solo, data warehouse es un viaje, no un destino.
- **Enfocarse en búsqueda de datos predefinidas y reportes periódicos:** Este es un error muy sutil, pero muy importante, el solucionarlo puede transformar a un administrador de data warehouse de librero de datos a héroe. La progresión natural de la información en el data warehouse es:
 - Extraer los datos de los sistemas históricos, limpiarlos y con ellos alimentar el data warehouse
 - Definir consultas predefinidas, hasta que se aprenda que es lo que los usuarios quieren
 - Convertir los reportes predefinidos en reportes de entrega regular

Esta es la progresión natural, pero no es la mejor, en ella se ignora el que los ejecutivos normalmente tiene su tiempo ocupado, y que los reportes son una obligación mas que un activo, a no ser que tengan tiempo para leerlos.

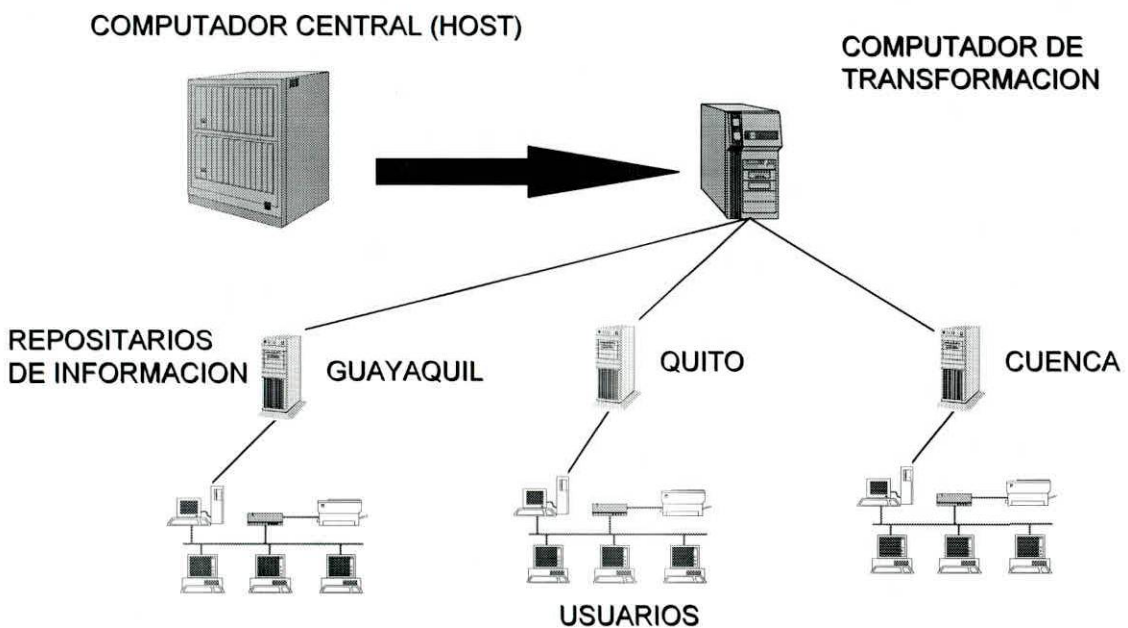
Sistemas de alerta, pueden resultar mas convenientes y pueden convertir al data warehouse en un sistema de misión crítica, estos sistemas monitorean el flujo de datos en el data warehouse a partir de un modelo predefinido, en el cual, el usuario establece sus factores críticos de control.

6. IMPLEMENTACIÓN DEL DATA WAREHOUSE

6.1. SELECCION DE HARDWARE

El proceso de selección de hardware en un proyecto de data warehouse, involucra un análisis de todos los niveles de la arquitectura propuesta, en cada uno de ellos, se debe establecer cuales serían los requisitos de hardware para poder implementar el proyecto. Esta selección puede ser tan sencilla como el requerir incrementar la capacidad en disco y la memoria principal en los computadores de los usuarios, como es el caso de un proyecto pequeño de data warehouse virtual, hasta requerir una estructura de hardware extremadamente compleja para cumplir con los objetivos del proyecto.

En la siguiente gráfica, se detalla la arquitectura propuesta para un banco, en esa arquitectura, era de suma importancia eliminar la dependencia de procesamiento en el computador central tanto por ambiente de programación como por tiempo de proceso.



En cada uno de los niveles se debe realizar un profundo análisis de las necesidades de hardware, en ciertos niveles como el de directorio de datos, su definición será sencilla y no implicará cambios durante el próximo año, sin embargo en los niveles de data warehouse y data mart, la selección de hardware será mas compleja y probablemente a los pocos meses se necesitará incrementar su capacidad o adquirir una configuración de mayor capacidad en cuanto a almacenamiento y velocidad de proceso, ante esto es importante cuando se plantea los requerimientos de compra de hardware, documentar la siguiente información:

- En que nivel se utilizará el recurso

- Capacidad de crecimiento de la configuración de hardware planteada
- Volumen de información que manejará
- Proyección de incremento de volumen de información, estableciendo periodos de tiempo en que se estime, que el hardware propuesto deba ser modificado o reemplazado
- Capacidad de reutilización del hardware en otros niveles del proyecto

Con esta información, el administrador del proyecto podrá justificar inversiones futuras, esto es importante ya que los volúmenes de información que se manejan en proyectos de data warehouse son muy variables, estos dependen de varios factores que son muy difíciles de estimar.

6.2. SELECCION DE SOFTWARE

6.2.1. BASES DE DATOS

Para el desarrollo del data warehouse, se puede utilizar dos tipos de bases de datos, estas son bases de datos relacionales y bases de datos multidimensionales.

Las bases de datos relacionales cumplen con el concepto de modelación relacional de datos, estas guardan los datos en tablas a través de registros con claves referenciales, siendo los datos accedidos a través de lenguajes comunes SQL (lenguaje de consulta estructurada), por otro lado las bases de datos multidimensionales, guardan los datos de una forma lógica en arreglos, debido a que no existe un modelo multidimensional estandar, no existe un método de acceso común. Algunos productos están disponibles como un motor de base de datos multidimensional, con sus propias herramientas de acceso.

En la actualidad existen bases de datos multidimensionales construidas a partir de este concepto, como es el caso de Essbase de Arbor software, también otras casas fabricantes de bases de datos como Oracle e IBM con su producto DB2, ofrecen addons a sus bases de datos para convertirlas en multidimensionales, en estas se puede combinar el guardar parte de los datos bajo un esquema relacional y la otra parte bajo un esquema multidimensional.

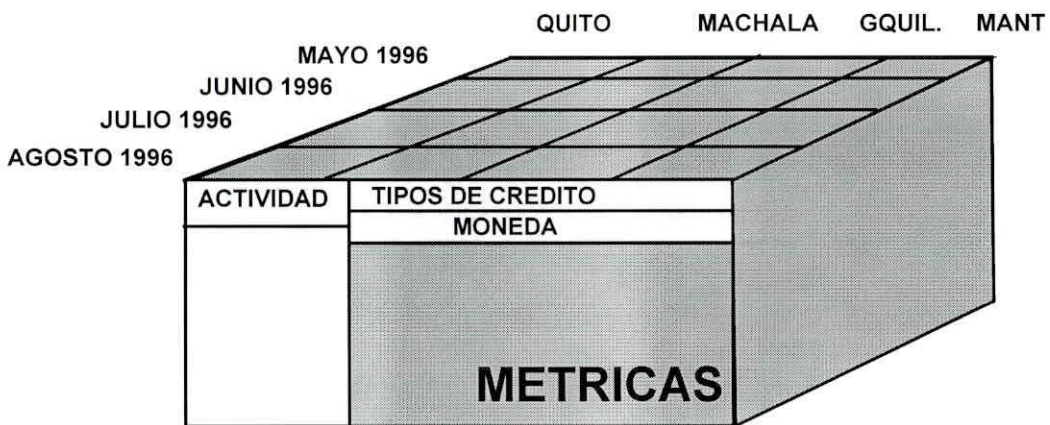
El siguiente cuadro detalla las diferencias entre bases de datos relacionales y multidimensionales.

	BASES DE DATOS RELACIONALES	BASES DE DATOS MULTIDIMENSIONALES
GRABACION INFORMACION, ACCESO Y VISTAS	<ul style="list-style-type: none"> • Relacional • Tablas de columnas y filas • Lenguaje SQL con extensiones • Herramientas de terceros usando api's 	<ul style="list-style-type: none"> • Dimensional • Arreglos: Hipercubo/Multicubo • Tecnología de matriz dispersa • Hoja de cálculo propietaria
USO Y EMPAQUETAMIENTO	<ul style="list-style-type: none"> • OLTP • Motor RDBMS • Búsqueda hacia abajo a un nivel de detalle • Perfomancia de consulta: variable 	<ul style="list-style-type: none"> • OLAP • Motor multidimensional • Búsqueda hacia abajo hacia niveles de sumalización y agregación • Perfomancia de consulta: rápida
TAMANO DE LA BASE Y ACTUALIZACIONES	<ul style="list-style-type: none"> • De gigabytes a terabytes • Almacenamiento de índices y denormalización, incrementa el tamaño • Consultas en paralelo y carga • Actualización mientras está en uso 	<ul style="list-style-type: none"> • Gigabytes • Compresión de datos y agregación dispersa • Difícil de actualizar cuando está en uso; cambios pequeños requieren reorganización.

Si bien es cierto, estas dos tecnologías permiten guardar los datos bajo diferentes esquemas, los cuales deben ser seleccionados de acuerdo a las características del proyecto, se debe tomar muy en cuenta que esto debe ser transparente para el usuario, para el la información debe estar archivada de cualquier forma, pero es responsabilidad del administrador del data warehouse presentarle al usuario una forma lógica de la manera como se debe ver esta información.

Una forma muy descriptiva de presentarle al usuario el almacenamiento de los datos, es a través del modelo de datos a partir de cubos, de esta manera los usuarios pueden usar términos de negocios que puedan entender, de una manera que les parecerá lógica.

La siguiente gráfica representa un cubo de información:

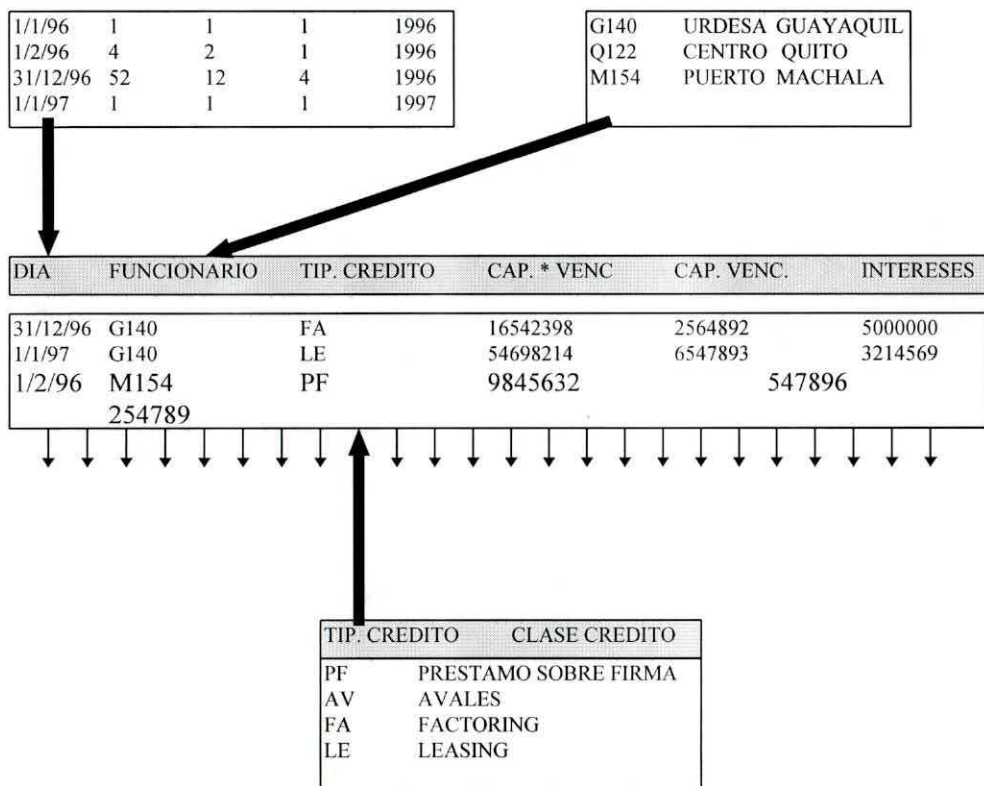


Este cubo, generado para la administración de la colocación de créditos en un banco, es la representación lógica de como se le presenta los datos al usuario final, la forma física como están guardados en la base de datos, es completamente distinta, sin embargo el objetivo es presentar un esquema de fácil entendimiento y similar al esquema mental de los usuarios. La intención es que se perciba al cubo como un objeto sólido al que se lo puede rebanar y obtener datos a partir de sus dimensiones,

los ejes dimensionales contienen métricas las cuales son los datos a analizar. Las variables que usamos para llegar a las métricas, es lo que llamamos dimensiones, en el gráfico anterior observamos cinco dimensiones: Mes, localidad, actividad, tipo de crédito y moneda, a través de los cuales se accesa a las métricas: Capital por vencer, capital vencido, intereses y comisiones.

El esquema del cubo, es la mejor manera de representar como las bases de datos multidimensionales intentan almacenar los datos, pero analicemos cual es la estrategia que se usa en las bases de datos relacionales.

En el siguiente gráfico se presenta el esquema estrella, comúnmente utilizado en las bases de datos relacionales.



En este gráfico donde se representa el esquema estrella, encontramos tres dimensiones: Fecha, funcionario y tipo de crédito, si analizamos cada una de ellas vemos que podemos pedir información altamente sumariada o al mas bajo nivel de detalle, por ejemplo si pedimos información de los valores por vencer de la Sucursal Guayaquil, durante el año 1996, recibiremos como resultado un registro donde se indica el total de colocación, pero si pedimos el total de intereses generados por funcionario y por tipo de producto recibiremos una gran cantidad de registros, a esto se le llama navegar a través de las dimensiones.

La decisión de que tipo de base de datos usar, dependerá de las características del proyecto, probablemente será conveniente usar los dos tipos en diferentes partes de la arquitectura del data warehouse.

Situaciones en que las bases de datos relacionales son la mejor opción:

- Para analistas de la estructura tope, la mejor opción es las bases de datos relacional cuando:
 - Requerimiento de análisis de alto nivel de agregados y/o datos muy detallados
 - El número de atributos dimensionales es muy grande
 - El número de valores únicos de un atributo dimensional es muy grande
 - Existen requerimiento inmediatos de definición dinámica de dimensiones jerárquicas
 - El propósito del datamart es el proveer un punto de vista corporativo en lugar del punto de vista de un analista específico.
- Para analistas orientados al análisis de los datos del negocio y al análisis de los procesos del negocio, bases de datos relacionales son su mejor opción.
- Bases de datos relacionales son mas eficientes en los procesos de refregamiento y carga de información, especialmente cuando hablamos de altos volúmenes.
- Las bases de datos relacionales soportan una mayor variedad de tipos de datos, como los de multimedia, fotos y sonido, los cuales podrían ser usados para análisis.

Situaciones en que las bases de datos multidimensionales pueden ser la mejor opción

- Para los analistas de la estructura tope, cuando realizan análisis de planificación, pronostico y presupuesto. El escenario de negocios para estas aplicaciones, normalmente realiza análisis de sensibilidad.
- Bases de datos mutidimensional son muy eficientes para cálculos matemáticos, que son necesarios en específicos datos analíticos derivados como varianzas, proporciones, distribución, etc.
- Este tipo de base de datos tiene un buen tiempo de respuesta, debido a que solo agregados precalculados son almacenados en las bases de datos multidimensional.
- Esta es la opción lógica cuando el control y administración del ambiente de base de datos va a ser realizado por el usuario.

6.2.2. HERRAMIENTAS DE EXTRACCION

La selección de herramientas de extracción, se la realiza en función de la configuración de hardware y software que tenga la empresa, adicionalmente se debe haber establecido los objetivos a lograr con el data warehouse.

Las herramientas de extracción de datos deben cumplir las siguientes funciones:

- Extraer información de diferentes ambientes de sistemas operativos y de bases de datos o sistemas de archivos
- Permitir la extracción de información usando un lenguaje de programación de uso común basado en SQL.
- Permitir la ejecución de programas de extracción en fechas y horarios establecidos.
- En algunos casos es importante que tenga capacidad de controlar información adicionada o modificada a partir de una fecha y replicación de información.

Existen diversas casas de software que han desarrollado sistemas que cumplen con una o varias de las funciones escritas anteriormente, la selección de estos productos, debe ser realizada cuidadosamente, ya que existen otros factores como tiempo de respuesta, sobrecarga en los servidores principales, etc. que deben ser analizados previo a su adquisición.

6.2.3. HERRAMIENTAS DE ANALISIS PARA USUARIO FINAL

El objetivo del data warehouse es crear datos para un sistema informacional, estos datos pueden ser revisados por los usuarios de diferentes maneras, reportes, cuadros estadísticos, gráficos de barras, etc. Para poder realizar este análisis, se le debe proveer al usuario de herramientas que le permitan acceder a esta información, estas podríamos clasificarlas de la siguiente manera:

- **Herramientas de extracción de información:** Sirven para extraer información de las bases del data warehouse, usando metodos sencillos sin tener conocimientos de programación, el objetivo es seleccionar bloques de información de acuerdo a un criterio de selección y ponerlos en una herramienta que les permita usarlos. Probablemente una de las herramientas mas usadas es el MSQUERY de microsoft, el que usando los filtros de ODBC (Open database conectivity), accesa a una gran variedad de bases de datos. Comunmente es usado desde Excell.
- **Herramientas de análisis:** Las herramientas de análisis normalmente tienen incorporado herramientas de extracción de datos, algunas lo hacen a través de filtros ODBC y otras traen sus propios filtros para las bases de datos más comunes. Estos productos están enfocados a brindar un ambiente de análisis al usuario, permitiendo desarrollar modelos en que se pueden usar funciones estadísticas, de decisión, financieras, matemáticas, etc. Entre las herramientas mas conocidas en el mercado están las hojas de cálculo como Excell, Lotus, Qpro, etc. y la herramientas de análisis estadístico como SPSS y SAS, algunas de estas herramientas dan facilidades para la construcción de sistemas de información ejecutivos.
- **Herramientas generadoras de reportes:** Estas herramientas están enfocadas a procesar de una forma ágil los bloques de información extraídos del data warehouse, su fortaleza principal, es brindar un ambiente de presentación de información muy facil de usar, a partir del cual usando funciones básicas como las estadísticas, matemáticas, de decisión, etc.y facilidades gráficas, el usuario pueda diseñar el formato en que desea presentar la información. Una de las herramientas mas conocidas en el medio es el IMPROMPTU, la cual da un ambiente muy amigable al usuario para extraer y presentar información.
- **Herramientas de análisis multidimensional:** Partiendo del concepto de análisis multidimensional, estas herramientas permiten analizar la información bajo diferentes criterios de acceso dimensional, por ejemplo las ventas totales de la empresa, pueden ser analizadas por producto y por fechas, o por zonas y vendedores o por fechas y vendedores. Para poder brindar esta facilidad de análisis multidimensional, la herramienta inicialmente debe definir un modelo de análisis, en el cual se configuran las dimensiones y las métricas, a partir de este

modelo, se extrae un bloque de información y la herramienta lo procesa generando un archivo bajo el concepto de multidimensionamiento. La forma especial en que es creado este archivo, es lo que permite que el análisis de la información sea rápida. En los casos en que las bases de datos sean multidimensionales, estas traen sus propias herramientas de análisis multidimensional, ya no necesitan generar estos archivos especiales. En el caso de la herramienta POWER PLAY, esta necesita crear estos archivos especiales, en el caso de la base de datos ORACLE, esta trae sus herramientas EXPRESS.

7. ADMINISTRACION DEL DATA WAREHOUSE

La administración del data warehouse, es un proceso muy complicado que implica el desarrollar las siguientes habilidades:

- El conocimiento del negocio en general, para poder comprender las necesidades del usuario.
- Conocimiento y desarrollo de tecnología que le permita agilizar la creación de soluciones.
- Conocimiento de gestión estratégica para poder orientar al usuario final.

Por otro lado es importante que en el proceso administrativo se tome en cuenta los siguientes factores:

- El data warehouse es una buena inversión solo si los usuarios pueden obtener información vital de una forma más rápida y barata que la que obtienen usando la tecnología actual.
- Es de mucha importancia el proveer medios que informen al usuario de la información que se está creando y cual podría ser su utilidad, el concepto de mercadeo del data warehouse es muy importante enfocarlo a la facilidad que se le da al usuario para que accese al directorio de datos.
- El mantenimiento de la estructura del data warehouse, es tan importante como cualquier otra aplicación de misión crítica, de hecho la experiencia indica que el data warehouse rápidamente se convierte en uno de los sistemas mas usados en la empresa.

Es importante tener en cuenta que los proyectos de data warehouse, en el largo plazo, crean nuevas demandas de sus sistemas operacionales; demandas de mejor calidad de datos, datos consistentes y nuevos tipos de datos.

7.1. *Análisis del negocio*

Existen tres formas en que se puede realizar el análisis del negocio bajo un punto de vista multidimensional: Análisis de la estructura tope, análisis de los datos de las transacciones del negocio y análisis de los procesos del negocio. Varios tipos de análisis del negocio, tienen las características de una o mas de estas formas.

Análisis de la estructura tope: Este tipo de análisis es el mas común, se enfoca principalmente a aprender acerca del negocio a través de usos intensivos de profundizar y enrollar los datos sumariados, datos detallados o a nivel de transacciones no se requieren para esta forma de análisis. Ejemplos de este tipo de análisis son: Análisis de ventas, análisis financiero, presupuestos y pronósticos.

Análisis de los datos de las transacciones del negocio: Este esquema, aprende del análisis de los datos que están asociados con las transacciones del negocio. No obstante los datos agregados pueden ser usados para identificar las transacciones del negocio, lo mas difícil de lograr en las tareas de análisis ocurre con datos detallados

sobre transacciones del negocio individuales, como por ejemplo el mercadeo de clientes personalizado, administración de microproductos, análisis de afinidad de productos. En estos ejemplos, es importante recalcar que la recabación de los datos sobre las transacciones del negocio, se obtienen a partir de los puntos de venta o de las ordenes de compra.

En este esquema es muy importante el análisis a nivel de detalle, la información sumariada es muy poco usada. En esta forma de análisis es importante el entender la naturaleza de las entidades y eventos, para de esta forma observar y relacionar entidades y eventos similares, podríamos determinar que uno de los usos mas frecuentes es el de establecer los perfiles de los clientes.

Análisis de procesos del negocio: Este tipo de análisis aprende sobre el negocio, a partir de los costos, ingresos y rentabilidad asociados con una cadena de eventos, para cualquier servicio o producto que se desee tomar decisiones sobre mejorar la calidad, eficiencia del negocio y niveles de servicio. Esta forma de análisis requiere que numerosos procesos (Ej: métodos de transporte), y segmentos (camino con inicio y fin), durante la ocurrencia del proceso, sean claramente definidos. La definición de datos es predominante en la tarea de implementación para estas soluciones. Ejemplos de este tipo de análisis son: Administración de producción, análisis de fuentes y canales, y análisis de precios para la industria del transporte.

Datos de las transacciones o datos agregados a un nivel muy bajo de detalle, son requeridos para el análisis del proceso de negocio debido a la dificultad en la definición de datos para segmentos, rutas y procesos. Los segmentos están compuestos de un punto de origen y un punto de finalización, entre los cuales sucede un proceso o servicio. Las métricas pueden ser el costo, volumen y apropiación de ingresos para los procesos dentro de un segmento, las rutas en este caso son agregados a los segmentos.

Segmentos, servicios y periodos son datos dimensionales para el análisis del proceso de negocio.

7.2. ADMINISTRACION DE USUARIOS

Una de las funciones mas complejas es la de administrar usuarios, partiendo del concepto de que data warehouse es la definición de una arquitectura para crear información que ayude en el proceso de toma de decisiones, podemos establecer que el negocio del administrador de data warehouse es brindar un servicio. Generalmente el cliente es parte de la empresa, sin embargo puede darse el caso de que se tengan clientes externos.

En un proyecto de data warehouse, generalmente al inicio se tienen pocos clientes y probablemente se conozcan sus necesidades, pero a medida que el data warehouse se desarrolle, la cantidad de clientes crecerá y resultará difícil tener un buen nivel de conocimientos de sus necesidades, este problema será mayor dependiendo del tamaño de la empresa, por lo tanto es importante definir una estrategia que nos permita administrar una gran cantidad de usuarios.

Existen tres etapas en que los usuarios se relacionan con el data warehouse:

- Conocimiento acerca del proyecto
- Participación en el proceso de creación de información
- Usuario continuo de información

En cada una de estas etapas, la estrategia de administración del usuario debe ser diferente.

Conocimiento acerca del proyecto: Este conocimiento el usuario lo puede adquirir de dos maneras: De forma informal o por presentaciones de la unidad de data warehouse, es importante tratar de que este primer contacto sea dirigido por quienes están desarrollando el proyecto, de otra manera se corre el peligro de una sobreventa de lo que es data warehouse, o peor aún de crear falsas expectativas en el cliente, quien al solicitar las supuestas maravillas del data warehouse y darse cuenta de lo que realmente es, sentirse frustrado y perder todo interés de usarlo.

En este primer contacto, es importante transmitir al usuario la siguiente información:

- Definir la misión y visión de la unidad de data warehouse
- Conceptos generales sobre la arquitectura propuesta
- Breve descripción sobre el análisis multidimensional
- Detallar el proceso de creación de información, haciendo hincapié en la importancia de la participación del usuario en este proceso

La intención de este primer contacto, es la de presentar al usuario una idea global de lo que es un proyecto de data warehouse, y por otro lado establecer por parte de el un compromiso en la creación de información.

Participación en el proceso de creación de información: El inicio del proceso de creación de información, generalmente parte de la necesidad puntual de un usuario, este plantea su requerimiento y en función de el se inicia el proceso de creación, sin embargo es conveniente que el administrador del data warehouse antes de aceptar requerimientos, realice un análisis de quienes serán sus clientes. Este proceso de conocer a sus clientes debe ser continuo, ya que inicialmente se enfocará a los usuarios que se encuentren en las áreas de negocio que abarque el proyecto de data warehouse, sin embargo conforme se vaya creciendo en otras áreas, se necesitará conocer a esos nuevos clientes, por otro lado es probable que en un año, un cliente haya cambiado, suprimido o agregado funciones que realiza.

La documentación de cada usuarios debe contener la siguiente información:

- Nombre
- Area en la que trabaja
- Cargo que desempeña
- Ubicación física
- Jefe inmediato
- Descripción detallada de sus funciones
- Detalle de los factores críticos de su gestión

Es importante en la documentación de las funciones y los factores críticos, documentar cada uno de ellos, tratando de crear funciones estándar, de tal manera que se pueda mapear a los usuarios con las funciones estándar definidas, por ejemplo

en un banco con varias sucursales con toda seguridad existirá mínimo dos gerentes de crédito, probablemente con las mismas funciones, si se clasifica sus funciones de una forma estandar, podremos a partir del requerimiento de uno de ellos, ofrecer al otro la misma información, o probablemente otras personas que ocupen otros cargos como el de analista de riesgo, realizarán una o mas funciones de los gerentes de crédito. Esto se complementa cuando a los cubos de información, les detallamos que funciones estándar cumplen.

Es importante tener en cuenta que existen dos tipos de usuarios del data warehouse, aquel que utiliza la información para controlar su gestión administrativa, como en el caso de los gerentes de crédito, para quienes es importante medir sus colocaciones y recuperaciones, este tipo de usuarios, usualmente definen sus necesidades al inicio, y posteriormente son esporádicos sus requerimientos, es más, son bastante predecibles en cuanto a sus necesidades de información.

Por otro lado están los usuarios cuyas necesidades de información es muy variada y generalmente cambian mucho la forma en que desean analizar los datos, este es el caso típico de los mercadotecnistas.

Usuarios continuo de información: Terminado el proceso de creación de información, generalmente los usuarios comienzan a hacer uso continuo de la misma, de acuerdo a las definiciones que se generaron como periodicidad de actualización, ellos esperan que los compromisos adquiridos por parte de la administración de información en cuanto a fechas de actualizaciones, se cumplan, esto es quizás uno de los factores que mas se debe cuidar, cuando el usuario llega a depender del análisis de información, retrasos en la entrega de la misma, pueden causar serios problemas.

La administración de usuarios es un factor crítico de éxito para la unidad de data warehouse, y se apoya mucho en la administración del metadata, por tanto es importante que el diseño del mismo contemple las necesidades de documentación descritas en este capítulo. Por último cabe destacar que la unidad de data warehouse debe cumplir el compromiso que tácitamente se establece con el cliente, el de brindar información oportuna y confiable.

7.3. ADMINISTRACION DEL METADATA

La administración del metadata, es probablemente uno de los factores mas críticos en un proyecto de data warehouse, el poder documentar todo el proceso de creación de información, será determinante para poder brindar un servicio eficiente.

En un proyecto de data warehouse, con un tiempo de vida de mas de seis meses, la probabilidad de que existan mas de cien cubos de información creados es muy alta, si analizamos el proceso de creación de un cubo, en el cual se debe establecer las necesidades del usuario, las reglas del negocio, los algoritmos de creación, los criterios de transformación, las fuentes de información, etc, va a ser muy difícil recordar como se realizó la construcción de esos cubos, si no tenemos la documentación del proceso, por otro lado al no disponer de esta documentación, va a

ser imposible distribuir a los usuarios las características de los cubos que forman parte del data warehouse.

Para comprender la administración del metadata, comencemos por analizar el concepto de metadata.

Metadata se define como los datos acerca de los datos, en una base de datos, metadata es la definición de los diversos objetos que comprenden una base de datos, en el caso de una base de datos relacional, esta representa la definición de las tablas, columnas, vistas, etc. En el data warehouse, el término metadata significa cualquier cosa que defina los objetos de un data warehouse.

Metadata abarca la definición de los siguientes tópicos:

- Localización y descripción de servers de data warehouse, bases de datos, tablas, nombres y sumalizaciones
- Reglas para la navegación automática hacia arriba, hacia abajo y a través de las dimensiones jerárquicas del negocio, tal como productos, mercados y cuentas.
- Nombres y alias definidos por el usuario para los campos nombrados con términos técnicos de difícil comprensión
- Reglas de cálculos definidas por el usuario
- Seguridad a personas y grupos, para la restricción al acceso, cambios sumalizaciones, cálculos y cualquier otro tipo de operación
- Descripción de fuentes originales y transformaciones
- Definiciones lógicas de tablas y atributos en el data warehouse
- Definiciones físicas de tablas y columnas, así como sus características
- Mapas de la interrelación de las tablas del data warehouse
- Historias de extracción
- Información de alias
- Algoritmos de sumalización
- Ubicación de área de negocio
- Historia de relaciones
- Dueño y conductor de los datos
- Patrones de acceso
- Tablas referenciales y datos codificados
- Criterio de archivo
- Indicador de calidad de datos
- Seguridad
- Unidades de medida

Adicionalmente, metadata también contiene componentes que permitan el mapeo de las siguientes tareas:

- Identificación de fuentes operacionales
- Cambiar de atributos simples a atributos mapeados
- Conversión de atributos
- Conversión de características físicas
- Conversión de tablas de referencia y codificación
- Cambios de nombre

- Cambios de llaves de acceso
- Uso y razones de valores por default
- Uso de algoritmos

Metadata es como un mapa de carretera para los datos, es muy similar a un catálogo de libros en una librería, en el cual describe el contenido del mismo y donde se encuentra físicamente ubicado, el metadata indica la localización y el contenido de diversa información contenida en el data warehouse.

El data warehouse debe tener un sistema que registre la información que contiene, este sistema debe ser organizado para brindar los siguientes servicios:

- Debe servir como un mapa para localizar donde se encuentra almacenada la información en el data warehouse.
- Es importante que el catalogo del data warehouse deba registrar dos grupos de información: Las definiciones necesarias para la tecnología de base de datos como nombres de tablas, dueño de la tabla, tipo de tabla, nombres de columnas, tipo de datos, etc. y las definiciones necesarias para el usuario del negocio. Esta definición de negocio, debe ser realizados en términos y definiciones de uso común en el negocio, y no deben estar sujeta a los límites impuestos por la tecnología. Como por ejemplo límites de longitud, caracteres no alfanuméricos, etc.
- Debe proveer una descripción por parte de los extractores que extraen datos de aplicaciones operacionales, y depositan los resultados en el data warehouse. Ellos son responsables de entender la organización de los datos que reciben y la organización de la base de datos donde los almacenan. Son responsables de los formatos de conversión, proveer valores por default, o truncar o redondear información.
- Debe almacenar las reglas del negocio con que se construye la información en el data warehouse. Reglas del negocio son la manera formal como las organizaciones resuelven inconsistencia de datos e implementan las políticas del negocio, también cuando los datos deben ser depurados en las bases de datos operacionales antes de guardarlos en el data warehouse.
- Debe guardar los controles de acceso y reglas de seguridad para permitir direccionar la administración de seguridad.
- El metadata debe registrar los cambios a través del tiempo. Es muy probable que los criterios usados en la recolección de datos hace dos años, actualmente sean distintos, estos cambios deben estar debidamente documentados.
- El metadata debe tener versiones para registrar sus cambios históricos.
- La estructura y contenido del data warehouse, se necesita tenerlo almacenado.
- El sistema de registro para el data warehouse, debe ser identificado de una forma clara y formal.
- La integración y transformación lógica que mueve los datos del ambiente operacional al del data warehouse, debe estar registrada en el metadata del data warehouse.
- La historia de actualizaciones debe estar registrada en el metadata, para permitir al usuario establecer el tiempo que abarca la información.

- Se necesita almacenar información que le permita al usuario calcular el tiempo que tomará su consulta.

El desarrollo del metadata, es de mucha importancia tanto en el desarrollo del data warehouse como en la operación posterior, vamos a analizar el impacto en las dos etapas.

Durante el desarrollo, el metadata es usado en los siguientes componentes:

- Extracción de datos fuentes
- Actualización de datos
- Transformación de datos fuentes
- Tortura de datos
- Agregación y sumarización de datos fuentes
- Diseño de la base de datos del data warehouse
- Diseño de consultas y reportes

Si el proceso de captura del metadata es realizado cuidadosamente durante estas actividades, el metadata estará automáticamente disponible para los usuarios finales cuando esté en producción el data warehouse.

El proceso de desarrollo, es un proceso de ingeniería, por lo tanto, debe ser documentado detalladamente, para permitir la difusión e incrementar la calidad del mismo, por lo tanto el metadata que resulte de este proceso, debe manejarse en función de versiones para controlar los cambios.

El enfoque principal en el diseño del metadata, es el de representar un amplio rango de relaciones al analista, frecuentemente con mucha redundancia, por otro lado debe proveer la documentación de todo el proceso a partir de los datos fuentes hasta los datos almacenados en el data warehouse, pasando por el proceso de extracción, refinamiento y re-ingeniería. Esta documentación debe ser mantenida para cumplir con lo siguiente:

- Verificación de calidad de datos: Documentación sobre los diversos cambios que han tenido los datos antes de almacenarlos en el data warehouse.
- Sincronización y actualizaciones: Conforme se genera nueva información, esta debe ser generada con el proceso de transformación definido cada vez que sea actualizada.
- Integración: Documentar las relaciones existentes entre los datos y las reglas del negocio de interés para el usuario.

A partir de que el data warehouse entre en producción, el metadata debe ser la herramienta que los clientes usarán para ubicar la información que les interese, adicionalmente información sobre reportes y consultas preestablecidas también pueden ser de interés para los usuarios.

Para el almacenamiento y administración del metadata, el administrador del data warehouse, debe tener muy en cuenta la funcionalidad que debe tener esta herramienta, sin embargo esta funcionalidad debe ser analizada de acuerdo a las necesidades de los dos tipo de usuarios de la misma.

Los usuarios técnicos de la herramienta, la usarán para la documentación de los procesos de creación, auditoría, verificación de calidad y consistencia de los datos, por otro lado los usuarios de negocio la utilizarán para analizar la existencia de información, consultas o reportes. La terminología usada por los dos tipos de usuarios, el tipo de información y su uso es diferente, por lo que es importante en el diseño analizar las diferentes necesidades de cada grupo.

Desde el punto de vista técnico, el metadata debe almacenar la siguiente información:

- Información sobre los datos fuentes, a partir de los que se crearon los datos en el data warehouse
- Reglas para la reestructuración de registros y campos, descodificación de valores de campos, reglas para el uso de valores en campos sin información y reglas para el chequeo de la integridad y consistencia de datos.
- Reglas para la transformación de datos, como en el caso de la transformación de datos para agregar el elemento de tiempo, definición de algoritmos para derivar valores y reglas para derivar sumalizaciones.
- Reglas para filtrar y mezclar datos de diferentes archivos y fuentes para alimentar el data warehouse.
- Metadata para datos objetivos, nombre y tipo de archivos, nombres de campos y sus alias, información sobre versiones, última fecha de actualización, frecuencia de actualización, campos indexados, relaciones, dueño de la información, usuarios autorizados, etc.

Desde el punto de vista de usuarios de negocio, les interesa la siguiente información:

- Descripción de los datos usando términos de negocio.
- Asociaciones entre nombres técnicos y los términos de negocio que los describen.
- Descripciones de los datos fuentes en términos de negocios y reglas de derivación descritas en términos comprensibles por el usuario final.
- Descripción de reportes y consultas predefinidas disponibles en el data warehouse, los términos usados, deben ser comprensibles por el usuario final.
- Información sobre los dueños de los datos.
- Requerimientos de autorización para el acceso a los datos.

El tomar en cuenta las necesidades de los dos tipos de usuarios durante el diseño del metadata, será determinante para que este pueda cumplir con sus objetivos.

7.4. Administración y selección de personal

En el desarrollo de un proyecto de data warehouse, es muy importante la definición del equipo de trabajo que trabajará en la construcción. Para poder establecer el equipo, es importante analizar las necesidades bajo dos criterios:

- El o los niveles de la arquitectura donde se desenvolverá la persona, dependiendo de esto se debe establecer el perfil necesario.
- Planificar las etapas del proyecto, en algunos niveles se necesitan personas que tengan habilidades especializadas por un corto plazo, tal como es el caso de la planificación e implementación del nivel de aplicaciones de mensajería, el cual en

el mediano plazo con toda seguridad no tendrá mayor modificación. Por otro lado otros niveles como el de administración de procesos en la fase de divulgación del producto, deben ser contratados de una forma permanente.

Ante esto es importante definir los siguientes puntos:

- ¿ Qué tipo de habilidades son requeridas para implementaciones de data warehouse ?
- ¿ Que recursos humanos existentes podemos utilizar en el proyecto ?
- ¿ Cual es el tiempo que se requerirá de las diferentes habilidades durante la ejecución del proyecto ?
- ¿ Que habilidades son requeridas solo durante la fase de desarrollo y que no se necesitarán durante la operación normal del data warehouse ?

Las habilidades necesarias y los tiempos de contratación de las personas que forman parte del proyecto, varía en función de la arquitectura definida.

Para establecer la cantidad de personas necesarias en un proyecto, primero se debe definir las funciones que se deben realizar, posteriormente de acuerdo a la carga de trabajo que tenga cada función, y el perfil necesario para poder cumplirlo, se determinará si una persona puede cumplir una o mas funciones, o una función necesita una o mas personas. Como ejemplo vamos a analizar el esquema de selección que se utilizó en un Banco.

Inicialmente se definieron las funciones que se debían realizar, para esto se definieron tres grupos de funciones: Técnicas, de relación con el usuario final y de apoyo corporativo.

Técnicas: Este grupo a su vez se subdividió en el grupo de desarrollo tecnológico y el de apoyo técnico.

El de desarrollo tecnológico tiene las siguientes funciones:

- Líder de proyecto: Administra el proyecto, obtiene recursos, participa en las sesiones de diseño.
- Analista de datos: Estudia el diccionario de datos de los sistemas operacionales, participa en la definición del modelo de datos del data warehouse.
- Analista de datos del negocio: Documenta los requerimientos de los usuarios finales y mantiene el metadata.
- Administrador de Base de datos: Crea físicamente las tablas de la base de datos, administra la perfomancia y realiza respaldos.
- Analista programador: Se encarga de desarrollar los programas de extracción/transformación y las aplicaciones de acceso de los usuarios finales.

El de soporte técnico tiene las siguientes funciones:

- Soporte técnico administrativo de sistemas: Determina el ambiente técnico, de redes, etc.
- Documentador técnico: Escribe y mantiene toda la documentación incluida en los manuales de usuario final, y documentación técnica.
- Capacitación: Crea los cursos de entrenamiento y entrena a los usuarios finales.

- Soporte a usuario: Soporta al usuario en el acceso al metadata y a los datos del data warehouse.

De relación con el usuario final: Este grupo de funciones está enfocado a quienes interactúan con el usuario final para determinar sus necesidades de información.

- Experto en el negocio: Desarrollar un profundo conocimiento del negocio de cada usuario a partir del análisis de sus necesidades.
- Analistas de necesidades de usuarios: Verifican modelos de datos, si existen los datos fuentes para crear la información solicitada y determinar la mejor forma de acceso por parte de los usuarios.

De apoyo corporativo: Estas funciones están encaminadas a los ejecutivos toques que deben ayudar al desarrollo del proyecto de data warehouse.

- Comité de finanzas: Apoyar en la decisión de inversión en los recursos necesarios para el desarrollo del data warehouse.
- Usuarios finales y de tecnología de información: Colaborar con el apoyo en cuanto a la metodología usada para establecer las necesidades de cada usuario, y por parte de los ejecutivos de sistemas, apoyar en las necesidades relacionadas con investigación, uso y adquisición de recursos tecnológicos.

A continuación se detallan las habilidades:

- Dominio de bases de datos
- Entender sobre bases de datos
- Dominio de modelación de bases de datos
- Entender sobre conceptos de data warehouse
- Dominio de análisis multidimensional
- Conocimientos básicos del negocio
- Conocimiento profundo del negocio
- Capacidad de comunicación con usuarios y personal técnico
- Habilidad de comunicación
- Objetividad, creatividad y diplomacia
- Trabajo en equipo
- Respeto y reconocimiento por parte de los ejecutivos de la empresa

	Lider de proyecto	Analista de datos	Analista de datos del negocio	Administrador de base de datos	Analista programador	Soporte tecnico administrativo de sistemas	Documentador técnico	Capacitador	Soporte a usuarios	Experto en el negocio	Analista de necesidades del usuario
· Dominio de bases de datos			X	X							
· Entender sobre bases de datos	X	X				X	X				
· Dominio de modelación de bases de datos		X	X	X	X	X					
· Entender sobre conceptos de data warehouse	X	X	X	X		X		X			
· Dominio de análisis multidimensional	X	X	X	X	X	X		X	X	X	X
· Conocimientos básicos del negocio	X		X					X			X
· Conocimiento profundo del negocio										X	
· Comunicación con usuarios y personal técnico	X		X					X	X	X	X
· Habilidad de comunicación	X							X		X	X
· Objetividad, creatividad y diplomacia	X									X	X
· Trabajo en equipo	X	X	X	X	X	X	X	X	X	X	X
· Respeto por parte de los ejecutivos de la empresa	X									X	

La matriz anterior, relaciona las funciones con las habilidades requeridas para ejercer esas funciones. Por último falta determinar la carga de trabajo que tendrá cada función, para de esta manera establecer la asignación de recursos humanos

Para la fase de diseño, no es mayor problema la definición del personal necesario, ya que se lo puede manejar a través de contratos por proyecto, sin embargo establecer cuantas personas se necesitan en la fase de producción, dependerá de la cantidad de requerimientos que reciba la unidad.

FUNCIÓN	# SEMANAL DE TRABAJOS	HORAS POR TRABAJO	TOTAL HORAS A LA SEMANA	HORAS POR RECURSO HUMANO	TOTAL DE RECURSOS
LIDER DE PROYECTO	NA	NA	NA	NA	1
ANALISTA DE DATOS	60	6	360	40	8
ANALISTA DE DATOS DEL NEGOCIO	80	2	160	40	4
ADMINISTRADOR DE BASES DE DATOS	NA	NA	NA	NA	1
ANALISTA PROGRAMADOR	50	4	200	40	5
SOPORTE TÉCNICO ADMINISTRATIVO DE SISTEMAS	20	2	40	40	1
DOCUMENTADOR TÉCNICO	30	4	120	40	3
CAPACITADOR	1	20	20	40	1
SOPORTE A USUARIOS	20	1.5	30	40	1
EXPERTO EN EL NEGOCIO	10	6	60	40	2
ANALISTA DE NECESIDADES DEL USUARIO	30	4	120	40	3

En función de este cuadro, se necesitaría 30 personas para la unidad de data warehouse.

8. CONCLUSIONES

El proceso de desarrollo de data warehouse, puede ser sencillo o complejo dependiendo de los objetivos y las circunstancias en que se realice, la necesidad de tener totalmente definido las metas que se desea lograr, es un requisito para lograr el éxito, sin embargo aquellos que piensan iniciar un proyecto de estos, deben estar conscientes de que este tiene fecha de inicio mas no de finalización. Data warehouse es un proceso continuo, donde constantemente se está creando información y realizando modificaciones en su arquitectura para brindar un mejor servicio.

La estrategia a seguir en el desarrollo de un proyecto de data warehouse, debe ser establecida a partir de un profundo estudio de las necesidades de la empresa y sus ejecutivos, así como también de su configuración de sistemas. Para lograr este análisis se debe seleccionar un grupo multidisciplinario de profesionales conformados por conocedores del negocio y expertos en tecnología de información, quienes en conjunto deben establecer la estrategia a seguir.

Por último quienes conformen la unidad administrativa del data warehouse, deben tener claro, que su negocio es el servicio de brindar y crear información, y que sus clientes en la mayoría de los casos, son clientes internos a quienes debemos dar la misma importancia que al cliente externo.

9. REFERENCIAS BIBLIOGRAFICAS

- Harjinder S. Gill & Prakash C. Rao, "The Official Guide to Data Warehousing", QUE
- Vidette Poe, "Building a Data Warehouse", Prentice Hall
- Lowell W. Fryman, "Data Warehousing: The Initial Project; Getting to Success", Intelligent Solutions.
- "Data Warehousing Concepts for AS/400", IBM
- W. H. Imon, "What is a Data Warehouse", Prism.
- "Multi-Dimensional Analysis: Extending the Information Warehouse Framework", IBM
- Lisa Houde and Michael J. Saylor, "True Relational OLAP", MicroStrategy
- Thierry Costa, "Three Issues in Data Warehousing", Hewlett Packard
- Peggy Imirie, "Your Data Warehouse: A Business Success or Science Project?", Tandem Computer
- Aaron Zornes, "A Taxonomy of Corporate Data Warehouses"
- "Multidimensional Database Oracle Express", Oracle