

Club de Investigación Tecnológica

Depósitos de datos

Preparado por: Beatriz Jiménez S., M.Sc. y Rafael Avalos C., M.Sc.

Noviembre 1998

**Editado y publicado por Rho-Sigma, S.A.
a nombre del Club de Investigación Tecnológica.
Todos los derechos reservados.
Prohibida la reproducción total o parcial.
San José, Costa Rica
Noviembre 1998**

Capítulo 1 Fundamentos de un depósito de datos

“Hoy los depósitos de datos (Data Warehouses) y los mercados de datos (Data Marts) constituyen el primer paso hacia lo que una vez fue el sueño de las bases de datos distribuidas ...”.

Alan Simon

Los depósitos de datos (DW) ofrecen excelentes posibilidades para la satisfacción de las necesidades de los usuarios finales encaminadas a la toma de decisiones. En busca de estas posibilidades, muchas compañías se apresuran en construirlos sin considerar el alto grado de dificultad en costo y tecnología para construirlos y mantenerlos, ni el impacto en las arquitecturas de sus sistemas, o cómo se integrarían con el resto de las aplicaciones.

El objetivo primordial de un DW es satisfacer las necesidades de información. Sin embargo, la tendencia ha sido centrarse en cómo aplicar la nueva tecnología, cuando la interrogante debería ser qué tan buena es dicha tecnología para solucionar los problemas empresariales focalizando la atención en sus objetivos y beneficios, más que en adoptar una tecnología por moderna y novedosa.

Según Colin White [White 1995] el objetivo de todo sistema de DW es incrementar la calidad y precisión de la información empresarial y entregarla en una forma accesible y comprensible. En su definición utiliza la palabra *sistema* estableciendo así la diferenciación de que un DW es simplemente un lugar físicamente separado donde almacenar datos, mientras que un *sistema de DW* proporciona una solución final completa para la entrega de información.

Un sistema de DW tiene dos componentes principales:

- Un almacén de datos transformados (depósito de datos, DW).
- Un conjunto de herramientas que apoyan el análisis estratégico de la información (sistemas de apoyo a la toma de decisiones, DSS).

El DW es un almacén de datos transformados e integrados que provienen de los sistemas operacionales o información externa, con una estructura diseñada para hacer más fácil y eficiente el acceso de los ejecutivos empresariales a la información requerida para análisis de tendencias y planificación estratégica.

En realidad, la tecnología incorporada en los ambientes DW no es nueva, es la convergencia de los Sistemas Administradores de Bases de Datos (SABD) que apoyan la toma de decisiones, la tecnología cliente/servidor que apoya el acceso distribuido a múltiples plataformas y la integración de herramientas de análisis a nivel de escritorio, que han servido de base al surgimiento de una estación de trabajo para el analista empresarial.

Según Alan Simon [Simon 1996] el interés que han despertado los DW no es más que una reacción al fracaso de los SABD Distribuidos, que no pudieron cumplir las promesas de sus desarrolladores de suministrar servicios como transparencia de localización, particionamiento de datos geográficamente, así como el manejo de la heterogeneidad. Detrás de todo esto se

encuentran los famosos “islotos de datos”, situación significativamente más problemática con el auge de las aplicaciones basadas en redes locales, las cuales conforman el escenario para la aparición de los sistemas DW.

Con la problemática descrita y la necesidad de extraer el máximo provecho de los datos distribuidos en toda la corporación, surgió un concepto aparentemente obvio: Si acceder a datos heterogéneos y distribuidos es demasiado complicado y problemático, *¿por qué no traer los datos necesarios a una determinada localización?*

De hecho, el DW es una versión “crecida” del proceso de extracción de datos que muchas organizaciones practicaron por muchos años como solución intermedia, que implementa definiciones del mundo de los SABD Distribuidos como réplica de datos, particionamiento, esquemas globales, protocolos de dos fases, etc., así como otros servicios de DW intermedios que automatizaban la extracción, calidad, movimiento y selección de datos entre diferentes ambientes.

Según [Inmon 1995] y [Dwop 1996] un DW es una colección de datos que apoya el proceso de toma de decisiones, caracterizado por estar orientado a temas, ser integrado, variable en el tiempo y no-volátil (no transitorio).

Orientado a temas

La información se clasifica en temas de interés. Para un empresario estos temas pudieran ser clientes, productos, contabilidad y vendedores; para una universidad los estudiantes, clases e instructores; para un hospital sus pacientes, equipo médico, medicinas, etc. Por ello, el DW organiza y orienta los datos desde la perspectiva de usuario final, cuyos almacenamientos están orientados a temas específicos.

En contraste, los datos operacionales están orientados a aplicaciones, por consiguiente dichos datos se organizan desde la perspectiva de la aplicación, de modo que el acceso de la aplicación a los datos tenga una mayor eficiencia. De esta manera, la información necesariamente no está organizada para que un analista - con herramientas gráficas de consultas inteligentes - pueda formular preguntas empresariales.

La diferencia radica en que los datos almacenados en los DW excluyen los datos que no serán usados para el proceso de toma de decisiones, mientras que los datos operacionales contienen datos que satisfacen los requerimientos funcionales y/o de procesamiento que pueden o no ser usados por el proceso de toma de decisiones.

Integración

Las organizaciones han administrado sus operaciones utilizando numerosas aplicaciones de software y múltiples bases de datos; por ello requieren de un DW para recopilar y organizar en un sólo lugar la información que sus aplicaciones han acumulado a través de los años.

En tal sentido, el DW se convierte el centro de la arquitectura de los sistemas de información que brinda la facilidad de integración en un mundo de aplicaciones informáticas aisladas, almacenando y organizando los datos necesarios para el procesamiento analítico de la información.

La esencia fundamental de un DW es que los datos que contiene están integrados. Esta integración que se encuentra dentro de un DW contrasta con la falta de integración en los entornos de aplicaciones, expresándose de diferentes formas como son: la consistencia en las convenciones para los nombres, las medidas de las variables, la codificación de las estructuras de datos y los atributos físicos de los datos, entre otros.

Independientemente de la forma asumida para diseñar el DW, el resultado debe ser consistente, ya que el ejecutivo para tomar las decisiones centra su análisis en la utilización de la información y no en dilucidar si el dato es o no confiable.

Variable en el tiempo

Típicamente en los sistemas operacionales los datos reflejan la situación actual, por lo que son constantemente actualizados y sobrescritos *in situ*, siendo poco usados los datos históricos. En contraste, la directiva empresarial requiere de un registro cronológico para análisis históricos encaminados a identificar y evaluar tendencias.

En este sentido, los DW se expanden en el tiempo al almacenar y manejar grandes volúmenes de datos históricos, con estructuras de datos que permiten su administración en el tiempo, siendo su alcance histórico y variable pues representan sus datos en un horizonte a largo plazo (5 a 10 años), mientras que el alcance de un sistema operacional es más a corto plazo (de la fecha actual hasta 60 a 90 días).

Los datos en un depósito son colecciones en el tiempo y son utilizados para comparaciones, análisis de tendencias y pronósticos, por ello su estructura de llaves contiene implícita o explícitamente un elemento de tiempo, tales como días, semanas, meses, etc. Ésta es otra manera en que se expresa la variabilidad en el tiempo de los DW.

No volátil (no transitorio)

Una vez que los datos son cargados en el depósito, sus registros no son actualizados, ya que constituyen una serie de instantáneas hechas a los datos en intervalos adecuados en el tiempo, suponiendo que fueron tomadas correctamente y en tal caso no hay que cambiarlas ni actualizarlas, siendo éste un recurso estable para garantizar la consistencia en informes y análisis comparativos.

Las instrucciones de actualización (inserción, modificación y eliminación) son tareas básicas en un ambiente operacional, pero en un DW como operación regular no hay actualizaciones, siendo la manipulación básica de los datos más simple: la carga inicial o poblacionamiento de los datos y el acceso de los usuarios a los datos.

Consecuencias derivadas de las operaciones dentro del DW

En el nivel de diseño: Al no ejecutarse actualizaciones, las precauciones para evitar anomalías de actualización no constituyen un factor importante en los DW, siendo más sencillo optimizar el acceso a los datos.

En el uso de la tecnología: La tecnología en ambientes operacionales debe tratar con aspectos complejos (transacciones, integridad de datos, recuperación después de fallas, detección y solución de interbloqueos) que son innecesarios para el procesamiento de los datos en un DW.

El diseño en un ambiente de DW es muy diferente al diseño operacional clásico, a pesar de que la fuente primaria de un DW son los datos que se encuentran en el entorno operacional. No obstante, es bastante común pensar que entre ambos entornos hay redundancia masiva, pero tal conclusión es superficial y denota falta de comprensión de lo que realmente ocurre dentro de un DW.

Los datos que llegan al DW provenientes de los entornos operacionales primero son filtrados - transformados e integrados - de manera tal que sólo llegan al DW los que tengan un valor real para los DSS. Además, desde una perspectiva cronológica hay muy poca superposición ya que los datos en los sistemas operacionales son muy recientes y los incluidos en el DW son más viejos, por tanto, no son los mismos que residen en los entornos operacionales.

Por otra parte, en un DW los datos no son cargados en tiempo real, sino que por el contrario, son refrescados de los sistemas operacionales en periodos regulares, de manera que la extracción y transferencia no afecte el rendimiento de los sistemas operacionales.

1.1 Usos y beneficios de un depósito de datos

Usos de un DW

El DW tiene tres usos primarios [Barquin 1996]:

1. Presentación de informes y gráficos en los formatos apropiados.
Si bien es cierto que las tareas de generación de informes han sido parte de las aplicaciones transaccionales tradicionales, el DW las amplía y consolida, al integrar los datos provenientes de los sistemas operacionales.
2. Apoya el análisis multidimensional.
Este tipo de análisis facilita la comparación de resultados, y simplifica la búsqueda de datos agregados a través de un número importante de características denominadas dimensiones.
3. Plataforma ideal para la minería de datos.
La minería de datos ayuda a descubrir de manera automática patrones de comportamiento en los datos, lo cual permite describir los datos existentes y predecir sus valores. En general, es la forma de responder a preguntas de las que no sabíamos lo suficiente como para cuestionárnoslas y que pueden ser fundamentales en un proceso de toma de decisiones.

Beneficios de un DW

Beneficios tangibles [Kimball 1996], [Berson 1996]

➤ El DW suministra *acceso* a los datos corporativos u organizacionales.

El DW constituye la tecnología que permite a gerentes y analistas acceder a los datos corporativos desde sus computadoras de escritorio, de manera inmediata y con altos rendimientos¹ de ejecución, con lo cual se incrementa su capacidad de autosuficiencia.

Este acceso también implica que las herramientas disponibles para los gerentes y analistas sean fáciles de usar, por ejemplo, que un determinado informe pueda ser ejecutado con el simple *click* de un botón una vez abierta la aplicación.

¹ Altos rendimientos significa que las consultas pequeñas deben ejecutarse en menos de un segundo.

- Los datos en el depósito son *consistentes*.

La consistencia significa que cuando dos personas independientes solicitan las ventas del Valle Central en enero de 1998, obtengan el mismo valor aún cuando la solicitud de los datos se ejecute en momentos diferentes.

La consistencia también significa que cuando se pregunte al depósito cuál es la definición de “ventas”, se obtenga una respuesta útil que permita encontrar lo que se está buscando en la base de datos (BD).

Finalmente, la consistencia también se expresa en que si los datos del día anterior no han sido completamente cargados, debe informarse al analista que los resultados no están disponibles debido a que el proceso de carga no ha finalizado.

- Los DW no son solo datos, sino también un conjunto de herramientas de consulta, análisis y presentación.

Los componentes del *back room*², el software de administración de la base de datos y los datos en sí, son solo cerca del 60% de lo que se necesita para un DW exitoso. El 40% restante lo constituyen las herramientas de consultas, análisis y presentación de datos.

- Los datos en el DW pueden estar divididos y combinados en todas las formas posibles. El requerimiento denominado *slice and dice*³ (dividir y combinar) hace referencia a la tecnología multidimensional, que facilita el desarrollo de las consultas y el análisis de los datos. Mediante este requerimiento se obtienen los encabezados de filas y las restricciones de las consultas que constituyen los bloques fundamentales de todo DW y que provienen de las dimensiones del modelo de datos.

- El depósito es un lugar donde se publican datos.

La responsabilidad de publicar es el centro de todo DW. Los datos no son simplemente acumulados en un punto central y dejados allí, sino que los datos son cuidadosamente integrados a partir de una diversidad de fuentes provenientes de toda organización, después – bajo la supervisión de un proceso de control de calidad - son limpiados y transformados, y sólo son dejados en el DW si cumplen los requerimientos de las solicitudes usuarias.

- Se mejora el proceso de rotación de inventario de los productos.
- Se disminuye el costo derivado por la introducción de un producto, pues se mejora el proceso de selección a través de campañas de mercadeo dirigidas.
- Se reduce la impresión y generación de informes.
- El proceso de toma de decisiones tiene un costo más efectivo, pues el procesamiento de las consultas *ad-hoc* no se realiza directamente en las bases de datos operacionales.

² Constituye el hardware central del DW.

³ Término común con que los analistas definen el análisis de los datos a lo largo de muchas dimensiones y a través de muchos subconjuntos.

Beneficios intangibles

- Se incrementa la productividad, al concentrar todos los datos requeridos en una única localización.
- Se mejora la relación con los clientes, ya que se incrementa el conocimiento de sus requerimientos individuales y tendencias, dada una relación más personalizada y una más rápida distribución de la información.
- Se activa el proceso de reingeniería. Los DW pueden proveer conocimiento útil acerca de los procesos empresariales, por lo que se pueden desarrollar ideas a través de las cuales se implementen estos procesos de reingeniería.
- Se incrementa la efectividad y rapidez en las decisiones empresariales, ya que los usuarios tienen acceso a amplia variedad de información multidimensional para realizar interesantes búsquedas y consultas.

El DW provee un conjunto de beneficios intangibles, que aun siendo más difíciles de cuantificar también se deben examinar cuando se planifica e implementa un DW.

1.2 Factores Críticos de Éxito

Para alcanzar el éxito en un proyecto de DW se deben identificar las áreas o procesos que tienen un fuerte impacto en los objetivos empresariales, para lo cual es recomendable tener en cuenta los siguientes aspectos [Sas 1997e]:

- Estado y condición de los objetivos.
- Listado de las áreas, procesos y unidades dentro de la organización.
- Listado de las áreas más estratégicas para la organización.
- Listado de los tópicos más sustantivos que mejor representan al área empresarial (productos, clientes, calidad de las ventas, mercadeo, etc.).
- Listado de los competidores, suministradores y compañías que han alcanzado la excelencia en esta área.

Una vez seleccionados estos elementos iniciales y a fin de asegurar el apoyo y los recursos necesarios es imprescindible:

- Obtener el compromiso de los patrocinadores de invertir en el proyecto.
- Escoger un equipo de trabajo que involucre tanto a funcionarios encargados del proceso de toma de decisiones como a especialistas del área de tecnología de información (TI), preferiblemente que tengan una visión amplia del funcionamiento de toda la empresa.
- Seleccionar los miembros más competentes y confiables que permitan una solución completa para el desarrollo del DW.
- Alcanzar la repercusión necesaria, para que los resultados parciales alcanzados puedan ser conocidos y probados.

No existe una fórmula que garantice el éxito en la construcción de un DW, sin embargo, la experiencia establece algunos aspectos que contribuyen al éxito [Inmon 1995]:

1. Establecer la correcta asociación entre los usuarios, la gerencia y el grupo informático.
Para lograr que los objetivos empresariales sean una realidad y asegurar que el DW contenga la información que satisfaga los requerimientos de la empresa, es esencial involucrar a los usuarios y a la gerencia. La gerencia puede ayudar a priorizar la implementación del DW y a seleccionar las herramientas de usuario final, mientras los usuarios justifican los costos del DW basados en sus expectativas y aspiraciones.
2. Seleccionar una aplicación piloto con alta probabilidad de éxito.
Con el propósito de definir si el DW es una tecnología básica para la empresa, es conveniente seleccionar una aplicación piloto de limitado alcance con un buen conjunto de herramientas medibles para los usuarios y la gerencia.
3. Construir prototipos con cierta frecuencia.
El diseño de un DW se fundamenta en un proceso iterativo de construcción de prototipos, una vez que se requiere de la retroalimentación con los usuarios para garantizar la satisfacción de los intereses empresariales.

En tal sentido, la única forma de asegurar que el DW satisfaga las necesidades de los usuarios es la generación de prototipos, a través de los cuales se exprese el proceso de implementación, demostrando con nuevos datos, información y nuevas características que el proyecto es imperativo para los intereses corporativos.
4. Implementación incremental.
Este tipo de implementación reduce el riesgo y asegura que el tamaño del proyecto se mantenga en rangos manejables.
5. Publicar historias exitosas.
La retroalimentación con los usuarios y la publicidad interna de cómo el DW los ha ayudado a trabajar más eficientemente ofrece una excelente oportunidad para dar a conocer sucesos exitosos dentro de la organización, lo cual puede contribuir favorablemente a la implementación del DW.

1.3 Razones para construir un depósito de datos

Existen fuerzas impulsoras que promueven la implementación de los DW y que determinan su necesidad, las cuales se encuentran en:

- El clima empresarial, caracterizado por un mercado global.
- Las cambiantes estructuras organizacionales, para garantizar la supervivencia y el éxito.
- Las tecnologías emergentes, que permiten su implementación.

Los DW surgieron con el afán de superar las limitaciones de los sistemas existentes para suministrar información de carácter sustantivo para las empresas; y se han convertido en un objetivo inmediato en el campo de la TI debido a diversas razones [Inmon 1992]:

Frustración de los sistemas legados

Una de las tareas más difíciles y frustrantes es realizar cambios significativos en los sistemas actuales, ya que no están preparados para asimilar tales transformaciones. Todo parece indicar que la única vía efectiva es la “cirugía” o la “extirpación”, pero muy pocas compañías están dispuestas a someterse a tan drástica medida, a pesar de que los DW les ofrecen la oportunidad de convertirse en proactivas.

Cuando una organización construye un DW y dirige hacia él toda la enorme carga que constituye el procesamiento analítico y los datos que apoyan a los sistemas de toma de decisiones, con ello liberan a los sistemas actuales de una gran presión, lo cual provoca un efecto de “limpieza” y agilidad que los hace mucho más efectivos.

Rendimiento

En los sistemas actuales se puede optimizar el procesamiento de transacciones, o flexibilizar el sistema para que responda mejor a la toma de decisiones, pero es muy difícil optimizar ambos a la vez. Sin embargo, con el desarrollo del DW se tienen dos ambientes: el operacional y el analítico, los cuales se pueden optimizar por separado y de este modo no entran en conflicto.

Administración de datos en el tiempo

Las aplicaciones actuales usualmente están diseñadas para almacenar la menor cantidad de datos posible en una aplicación, ya que mientras mayor sea la cantidad de datos que se tenga, mayor será la dificultad para organizarla, reestructurarla, indexarla y monitorearla, pero su horizonte es limitado en el tiempo.

Por su parte, la administración de las organizaciones necesita de una perspectiva a largo plazo para detectar tendencias y obtener una visión más amplia, para lo cual los DW - con su perspectiva de largo plazo - ofrecen una nueva y única posibilidad.

Apoyo a las estrategias empresariales

El DW es un mecanismo indispensable para la administración desde el punto de vista:

Empresarial

- Impulsa la creatividad de los encargados de la toma de decisiones ejecutivas con el propósito de incrementar la competitividad y asegurar una adecuada rentabilidad.
- Suministra datos integrados y resumidos a través de aplicaciones y herramientas de usuario final.
- Facilita el proceso de análisis, pues permite analizar la información de diversas maneras para desenterrar hechos acerca de clientes, mercados y productos, y utilizarla como una posible ventaja competitiva.
- Identifica y evalúa tendencias a largo plazo en función de sus datos históricos.

Tecnológico

Las TI están cambiando rápidamente, y sus capacidades se están incrementando, como evidencian los siguientes aspectos:

- El precio del procesamiento computacional medido en MIPS (millones de instrucciones por segundo) continúa decayendo, mientras que el poder de los microprocesadores se duplica casi anualmente.
- El precio del almacenamiento digital continúa decayendo.
- Los anchos de banda de las redes de comunicación se siguen elevando, mientras que su precio no se incrementa en la misma medida, aunque se mantienen relativamente altos.
- El legado de los sistemas actuales puede ser integrado con nuevas aplicaciones.

Apoyo a la arquitectura cliente/servidor

Los DW pueden verse como un *servidor* gigantesco que suministra datos precisos e integrados de múltiples fuentes, y los *clientes* como nodos que acceden a dicho servidor.

Los DW expanden el alcance de la información ya que sus usuarios no son profesionales de computación sino expertos en el negocio que tienen acceso a amplia información presentada consistentemente y disponible a través de sus estaciones de trabajo.

Además, el DW al crear una fuente única de la “verdad” permite una distribución más rápida de la información a los usuarios, la reducción de la demanda de consultas personalizadas, y la reducción en la impresión y generación de informes⁴.

No obstante todas estas razones, cuando se trate de mejorar y aumentar el acceso a las fuentes de información de la organización, habrá personas a favor o en contra, por ello estos argumentos – considerando el poder de la información - deben estar bien fundamentados.

1.4 Pautas por considerar en el desarrollo de un depósito de datos

Muchas compañías han emprendido el camino de desarrollar DW que les suministren información precisa para la toma de decisiones más acertadas. Sin embargo, para incrementar de forma significativa las oportunidades de éxito en el desarrollo e implementación de un DW, deben examinar las siguientes pautas [Poe 1994]:

1. Definir claramente el objetivo.

La pregunta primordial es si el DW es parte central del plan estratégico de la empresa y si determina un cambio de arquitectura y plataforma. La tarea consiste en fundamentar *¿por qué se desea un DW?*, además el beneficio esperado debe ser detallado y documentado.

2. Comprender cómo el DW se ajusta a la arquitectura de los sistemas empresariales.

En general, toda nueva tecnología trae consigo la necesidad de administrar la arquitectura y los sistemas de información de una manera más sofisticada, ya que se requiere de nuevas formas de comprensión sobre cómo los datos funcionan dentro de la arquitectura de los sistemas en funcionamiento, considerando aspectos relacionados con redes de comunicación, plataforma, interfaces de usuarios, herramientas de consulta, etc.

⁴ Los elementos de datos están listos y accesibles para otros usos, no solamente para un tipo de informe en particular. Además, se minimizan los errores que ocurren durante la preparación de reportes.

Cuando se comienza a desarrollar un DW, han de analizarse interrogantes como:

- ¿El DW será una BD centralizada?
- ¿Su sistema hace más fácil el acceso de datos a los usuarios?
- ¿Nos encontramos ante un sistema de datos distribuidos?
- ¿Puede la configuración de red administrar el incremento de los flujos de datos?
- ¿Está en condiciones el SABD de soportar el esperado incremento de datos para el desarrollo del DW, sin una degradación en su rendimiento?

3. No pensar que un DW puede desarrollarse en un corto período de tiempo.

La consolidación de los datos de diversas fuentes hacia una estructura y aplicación específica es sólo el comienzo, ya que para extraer y transformar los datos en información, se deben superar otros problemas:

- Datos redundantes en diferentes sistemas con diferentes valores.
- Datos que son cambiados por la estación de trabajo de un usuario que alimenta a otro sistema, de manera que los datos originales difieren de los reportados.
- Campos de datos que almacenan información que debe estar en campos separados.

Muchas compañías fracasan en su primer intento por desarrollar un DW debido a que subestiman la magnitud del análisis de datos y la complejidad que significa trabajar simultáneamente con diferentes nuevas tecnologías, lo cual exige capacitar al equipo de trabajo y a los gerentes encargados de la toma de decisiones.

La comprensión de estas complejidades involucradas en el análisis, extracción, transformación y consolidación de los datos de los sistemas actuales, permite dilucidar la cantidad de tiempo y recursos necesarios para el desarrollo de un proyecto de DW.

4. Identificar claramente las responsabilidades del equipo de trabajo.

En la medida que se conozca el alcance de lo que hará el equipo de trabajo, más sencillo será el desarrollo del proyecto de DW. Entre las tareas que el equipo de trabajo debe ejecutar una vez delimitadas sus responsabilidades, se encuentran:

- Modelar y analizar los datos.
- Definir requerimientos.
- Definir fuentes de datos.
- Reconciliar posibles conflictos de datos.
- Programar el sistema para poblar el DW a partir de los sistemas existentes.
- Investigar qué herramientas de usuario final se utilizarán.
- Seleccionar el equipo para capacitar a los usuarios finales.

5. Comprender la diferencia entre datos operacionales y datos de apoyo a las decisiones.

Los datos operacionales son usados por los sistemas funcionales empresariales, mientras que los datos de apoyo a las decisiones proporcionan la información necesaria para el análisis de tendencias y planificación estratégica por parte de los ejecutivos.

Las diferencias entre estos dos tipos de datos se resumen en el cuadro No. 1-**Error! Unknown switch argument.**

Cuadro No. 1-1 Diferencias entre datos operacionales y de apoyo a la toma de decisiones

Datos operacionales	Datos de apoyo a las decisiones
Actuales	Actuales + Históricos
Cambian continuamente	Estable
Datos diarios	Datos históricos
Detallado	Integrado y resumido
Pueden ser actualizados	No son actualizables
Orientado a aplicaciones	Orientado a temas
Diseño de BD no redundante	Puede haber redundancia
Muchos usuarios de forma concurrente	Pocos usuarios de forma concurrente
Actualizaciones y consultas predefinidas	Consultas complejas y no predefinidas
Pequeñas cantidades de datos detallados	Grandes cantidades de datos detallados
Requerimiento de respuesta inmediata	No es factor crítico el tiempo de respuesta

6. Planificar un presupuesto apropiado para las herramientas que ayudan en las etapas más delicadas del desarrollo del DW.

La mayor dificultad se concentra en las siguientes etapas del ciclo de vida:

- El análisis de datos para buscar las mejores fuentes de datos de los sistemas actuales.
- La programación requerida para transformar, limpiar, integrar y resumir datos que puedan estar en diferentes BD y plataformas.

No se debe subestimar este esfuerzo, por ello están apareciendo en el mercado herramientas (*middleware*⁵) que automatizan este proceso y generan el código para extraer, transformar, generar y cargar datos, proporcionando funcionalidad; tales herramientas se deben considerar como parte del presupuesto.

7. Escoger herramientas de *front-end*⁶, de consulta y de acceso a los datos basadas en las necesidades y habilidades de los usuarios.

Muchas de las herramientas de consulta son extremadamente sofisticadas y están dirigidas hacia el desarrollo de aplicaciones que requieren gran conocimiento del modelo relacional de base de datos; distintivamente, un requerimiento esencial de un DW es que las herramientas de *front-end* y de consultas se ajusten a las necesidades, habilidades y nivel de confort de los usuarios finales.

La selección de las herramientas de consulta puede afectar significativamente la forma en que el DW será recibido por la comunidad usuaria, la cual estará ansiosa o escéptica de sus beneficios y por ende de su utilización, en dependencia de la curva de aprendizaje y del valor agregado real.

⁵ Término genérico que se utiliza para definir el conjunto de software, servicios y herramientas de hardware que intercambia información en forma transparente entre aplicaciones y bases de datos, donde se ocultan para los *clientes* los elementos específicos de cómo funcionan los entornos de comunicación y los mecanismos de acceso a los datos en los *servidores* [Berson 1997].

⁶ Herramientas de usuario final que se clasifican en aplicaciones y herramientas. Las aplicaciones son desarrollos particulares de software del tipo llave en mano. Las herramientas incluyen compiladores, lenguajes de programación, generadores de reportes y otros [Hernández 1997].

8. Obtener la experiencia necesaria.

Las empresas que han implementado satisfactoriamente sus DW, conforman de manera ideal un equipo de trabajo con las siguientes habilidades:

- Análisis, diseño y administración de BD.
- Experiencia con el sistema actual.
- Análisis empresarial y trabajo con usuarios.
- Programación y diseño de interfaces gráficas.

Muchas prácticas y reglas tradicionales de la TI son desechadas en el DW; por ello, no es aconsejable esperar a que los analistas conozcan cómo construir el depósito, sino que es preciso capacitar a los analistas y capacitar a los usuarios.

La comprensión de la importancia de analistas capacitados, equivale a mejores toma de decisiones y a la reducción de errores. Por su parte, los usuarios no sólo deben conocer cómo utilizar el depósito sino que deben ser capaces de transmitir este conocimiento a otros. De la capacidad y el entrenamiento técnico recibido, se obtienen las habilidades necesarias para garantizar que este esfuerzo sea exitoso.

1.5 Mercados de datos

Los DW contienen información detallada así como datos integrados e históricos que son comunes a toda la organización, los cuales pueden suplir las necesidades de los diferentes departamentos en sus procesamientos para los DSS, mientras no contengan muchos datos.

Sin embargo, los DW no se mantienen pequeños durante mucho tiempo, en consecuencia:

- La competencia dentro del depósito se intensifica, debido a que cada vez más departamentos tratan de utilizar el DW para el procesamiento de sus DSS, lo cual se convierte en un problema al tratar de satisfacer los requerimientos de todos los departamentos que lo están accediendo en un momento específico.
- El proceso de satisfacción de consultas de manera personalizada es cada vez más difícil. En un DW con pequeñas cantidades de datos, el analista de los DSS puede producir análisis personalizados y agregados, pero en el caso de grandes volúmenes de datos dentro del DW, el analista no tendría tiempo ni recursos para elaborar tales análisis personalizados.
- A medida que se incrementa el volumen de datos, el costo del procesamiento dentro del depósito se incrementa.
- El software que está disponible para el acceso y análisis de grandes cantidades de datos (típico en los DW) no es tan refinado como el software para pequeñas cantidades de datos (típico en los *mercados de datos*).

Los DW están propensos a contener grandes volúmenes de datos, de los cuales no todos son de interés para todos los usuarios. Para administrar esta situación, se seleccionan los datos cuyo rango abarque a toda la organización y se dividen en subconjuntos que satisfagan concretamente a departamentos, aplicaciones o funciones empresariales.

Los mercados de datos son una extensión natural de los DW; son BD que contienen información de interés exclusivo para un determinado sector de la empresa, que apoyan a los DSS departamentales con una arquitectura que se fundamenta en la arquitectura de los DW.

Los datos que residen en un DW están a un nivel muy granular⁷ y los que residen en el mercado de datos poseen la misma característica pero con un nivel más refinado, ya que contienen diferentes combinaciones y selecciones de los datos detallados que se encuentran en el DW [Inmon 1996a].

El usuario de los mercados de datos es un analista departamental de los DSS que realiza el proceso de toma de decisiones desde el punto de vista departamental. Estos analistas encuentran muy atractivos los mercados de datos, ya que cuando un departamento tiene su propio mercado de datos:

- Puede personalizar con mayor nivel de detalle los datos; es decir, puede resumir, ordenar y estructurar sus propios datos sin necesidad de considerar otros departamentos, y sin que los datos en el mercado tengan que satisfacer todos los requerimientos de la organización.
- Puede seleccionar una cantidad mucho menor de datos históricos que la que almacena el DW, ya que la cantidad de datos históricos está en función del departamento y no de toda la corporación.
- Puede ejecutar sus procesamientos de la forma que deseen y cuando lo deseen, sin considerar para la utilización de los recursos a otros departamentos, pues los datos que accesan y procesan son referidos a sus departamentos.
- Puede seleccionar el software que mejor se ajuste a sus necesidades, ya que existe una mayor riqueza de software a nivel de mercados de datos.
- El costo unitario de almacenamiento y procesamiento en los mercados de datos es muy inferior al utilizado por los DW.

Con el crecimiento del DW, el mercado de datos se convierte en un fenómeno cada vez más atractivo debido a las diversas razones económicas, organizacionales y tecnológicas que hacen tan apetecible su construcción.

⁷ Término que se utiliza en los DW para expresar el nivel de detalle. A mayor nivel de granularidad, mayor nivel de agregación y por tanto menor nivel de detalle.

Capítulo 2 Arquitecturas de depósitos de datos

Factores críticos en el diseño arquitectónico

La fundamentación arquitectónica debe basarse en los requerimientos del proceso de extracción y transformación de los datos destinados a la BD del depósito y facilitar el análisis empresarial con un acceso transparente a la BD. La especificación de una arquitectura de DW requiere de tres importantes elementos:

➤ La BD del DW.

El dilema es si la BD se debe manipular por medio de un Sistema Administrador de Bases de Datos Relacional (SABDR) o por uno propietario conocido como SABD Multidimensional (SABDM).

➤ Las técnicas de extracción.

El objetivo es determinar si los datos administrados por el depósito utilizarán las facilidades de los SABD o utilizarán técnicas especializadas de extracción de datos.

➤ Las herramientas de front end.

La pregunta está referida a si los datos accedidos desde el DW emplearán consultas relacionales, vistas multidimensionales de datos relacionales, interfaces personalizadas o de propósito especial para el proceso de toma de decisiones.

Bases de datos relacionales

Muchas organizaciones no tienen un fuerte requerimiento operacional que las obligue a utilizar sistemas propietarios para garantizar las funciones de un depósito, ya que en muchos casos el DW puede construirse utilizando un SABDR y las herramientas de propósito general para el acceso a los datos de que disponible la empresa.

Las arquitecturas de los DW basadas en SABDR funcionan como repositorios centralizados de datos y casi todos los SABDR ofrecen una respuesta relativamente rápida para las consultas formuladas por una gran cantidad de usuarios conectados en línea. Estos SABDR brindan servicios transparentes a través de las herramientas de consultas, plataformas cliente/servidor e interfaz SQL. Las últimas versiones de productos de SABDR vienen preparadas para satisfacer consultas complejas y funciones de análisis, tanto para los datos operacionales como para el depósito.

Algunos de los proveedores que suministran productos con funcionalidad especializada para el depósito son [Barquin 1997]:

- CA-Open Ingres de Computer Associates.
- Data Warehouse de Information Builders.
- Informix OnLine de Informix.
- Oracle 7.x de Oracle Corp.
- WORKS de Sybase.
- Visual Warehouse de IBM.

Estas herramientas procesan las BD bajo sistemas en paralelo, incluyendo consultas, índices, cargas, servicios de respaldo y de recuperación. Las consultas son controladas por mecanismos de particionamiento dinámico que permiten múltiples procesadores en paralelo.

Bases de datos propietarias

Estos sistemas son importantes cuando se desea satisfacer de manera prioritaria algunas aplicaciones de apoyo al proceso de toma de decisiones, tales como el análisis de bases de datos multidimensionales para descubrir tendencias de mercado y evaluar la efectividad de las campañas promocionales. Para estas aplicaciones una tecnología propietaria está plenamente justificada debido a que se basan en un procesamiento multidimensional.

Los proveedores de herramientas para soluciones de DW propietarias, plantean que los SABDR convencionales, no son capaces de suministrar altos rendimientos en ambientes de apoyo al proceso de toma de decisiones debido a que son optimizados para cubrir altas velocidades en los procesamientos transaccionales en línea. En contraste, los sistemas propietarios de DW incorporan una plataforma diseñada para responder más rápidamente a las consultas en línea. Además, estos sistemas satisfacen otras características como:

- Segmentación y particionamiento de archivos.
- Procesos de almacenamiento.
- Personalización de funciones para el desarrollo del DW.
- Consulta iterativa.
- Ampliación del SQL para alcanzar altos rendimientos en el DW.

Por otra parte, las herramientas propietarias tienen significativas limitaciones que incluyen: degradación del rendimiento en la medida que la BD se incrementa y baja tasa de carga del DW. Otro aspecto importante es que las bases de datos multidimensionales no son compatibles con el SQL estándar y no se integran bien con los SABDR convencionales.

No obstante, las soluciones propietarias son utilizadas en muchas organizaciones debido a las características de su solución multidimensional de apoyo a los OLAP y a la implementación de mercados de datos dirigidos a satisfacer necesidades departamentales.

En el presente capítulo se incursiona por diferentes estructuras arquitectónicas para DW, que hemos denominado:

- Arquitectura básica.
- Arquitectura genérica.
- Arquitectura por niveles.
- Arquitectura por bloques.
- Arquitectura basada en el modelo empresarial.

2.1 Arquitectura básica

Un DW tiene una arquitectura bien distintiva, conformada a partir de diferentes niveles de agregación y detalle, como se muestra en la Figura No. 1 [Inmon 1995]. Estos diferentes niveles de datos dentro de un DW reciben diferentes niveles de utilización; como regla, a mayor nivel de agregación mayor nivel de utilización.

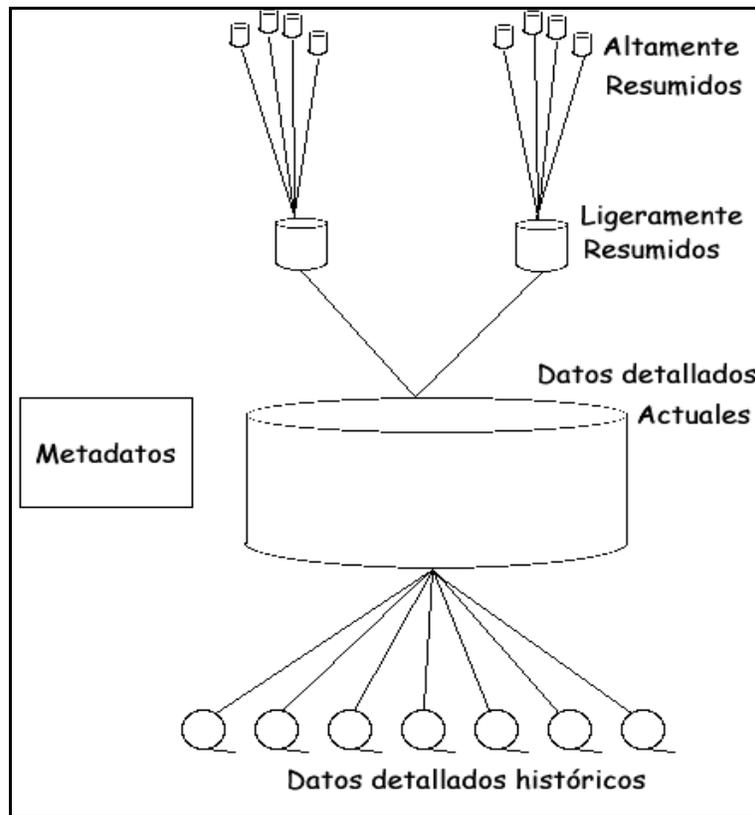


Figura No. **Error! Unknown switch argument.** Arquitectura de un depósito de datos

Los datos en los niveles más altos de agregación se pueden indexar y reestructurar con relativa facilidad, mientras que los datos con mayor nivel de detalle debido a su volumen, no se pueden indexar y estructurar fácilmente.

Los accesos a los datos más resumidos son rápidos, más eficientes y necesitan menos recursos que los datos con un mayor nivel de detalle; bajo este enfoque, se puede garantizar una mejor utilización de los recursos.

Los niveles principales de resumen son:

Datos actuales detallados

Son los eventos más recientes de gran utilidad que son almacenados al máximo nivel de detalle, por lo tanto requieren de grandes volúmenes de almacenamiento. Estos datos usualmente se almacenan en discos muy rápidos pero caros y complejos de administrar.

Datos históricos detallados

Poseen el mismo nivel de detalle que el anterior, almacenados en algún dispositivo de almacenamiento masivo. A este tipo de dato no se accede frecuentemente, por ello no es obligatoria la utilización de discos, sino de dispositivos de almacenamiento más pequeños.

Datos ligeramente resumidos

Son datos filtrados de los niveles detallados, generalmente almacenados en discos. En este nivel arquitectónico los elementos de diseño por considerar son:

- En qué unidad de tiempo se realizarán las agregaciones o resúmenes.
- Cuáles serán los contenidos (atributos) de datos que serán resumidos.

Datos altamente resumidos

Estos datos se caracterizan por ser fácilmente accesibles y por estar bien compactados. Esta información se almacena en algunos casos dentro de la propia estructura del DW y en otros casos puede residir de forma física fuera de su estructura, e incluso así este tipo de dato forma parte de la arquitectura de un DW.

2.2 Arquitectura genérica

Para minimizar el impacto de los requerimientos de los DSS en el rendimiento de los sistemas operacionales y debido a que las aplicaciones de los DSS son totalmente diferentes en naturaleza de las aplicaciones operacionales, es de vital importancia preparar los datos que satisfagan estos nuevos requerimientos, creando un DW como repositorio de información separado, usado estrictamente para los propósitos de los DSS.

Para puntualizar presentamos la distinción entre sistemas operacionales y sistemas de apoyo a la toma de decisiones (DSS) [Kulkarni 1997]:

Los sistemas operacionales son:

- Vitales para la eficiencia diaria de la organización.
- Usados para automatizar una determinada rutina y tareas de tipo predecible.
- Caracterizados por grandes volúmenes de pequeñas transacciones, que generalmente están limitadas en alcance.
- Diseñados para obtener *datos*.

Los sistemas de apoyo a la toma de decisiones (DSS) son:

- Vitales en la generación de ventajas competitivas que impactan fuertemente en la rentabilidad a largo plazo de las empresas.
- Construidos para activar la exploración, análisis y presentación de la información incluyendo aspectos de carácter investigativo que no tienen un aspecto predecible.
- Caracterizados por pocas preguntas, las cuales por lo general son de amplio alcance.
- Diseñados para obtener *información*.

Por tales razones, los datos almacenados en el DW necesitan un formato estructurado, integrado y consistente, aún cuando las fuentes de donde provengan almacenen sus datos en formatos diferentes, por ello la información cargada para el DW debe ser primero convertida en el formato establecido por el DW.

Esta integración de datos constituye un aspecto clave de los beneficios que reporta un DW, pero suele subestimarse el considerable esfuerzo que ocasiona dicha integración cuando existen diferentes fuentes de datos derivadas de múltiples transacciones y se necesita incorporarlas, de manera consistente, en el DW.

Los datos administrados por los sistemas operacionales constituyen la fuente principal para los DW (ver Figura No. 2) [Harjinder 1996] [Berson 1997], pero antes de introducir los datos en el DW se deben resolver las inconsistencias, ya que las bases de datos operacionales están diseñadas para apoyar aplicaciones en variados formatos.

En tal sentido, los DW surgen como una herramienta de gran importancia que permiten:

- Extraer u obtener los datos de múltiples fuentes, limpiarlos y reformatearlos.
- Reconciliar diferencias en la semántica de los datos, fechas de transacciones, diferencias monetarias, etc.
- Consolidar los datos en mayores niveles de agregación.
- Responder a interrogantes empresariales mediante el uso de datos externos e internos.

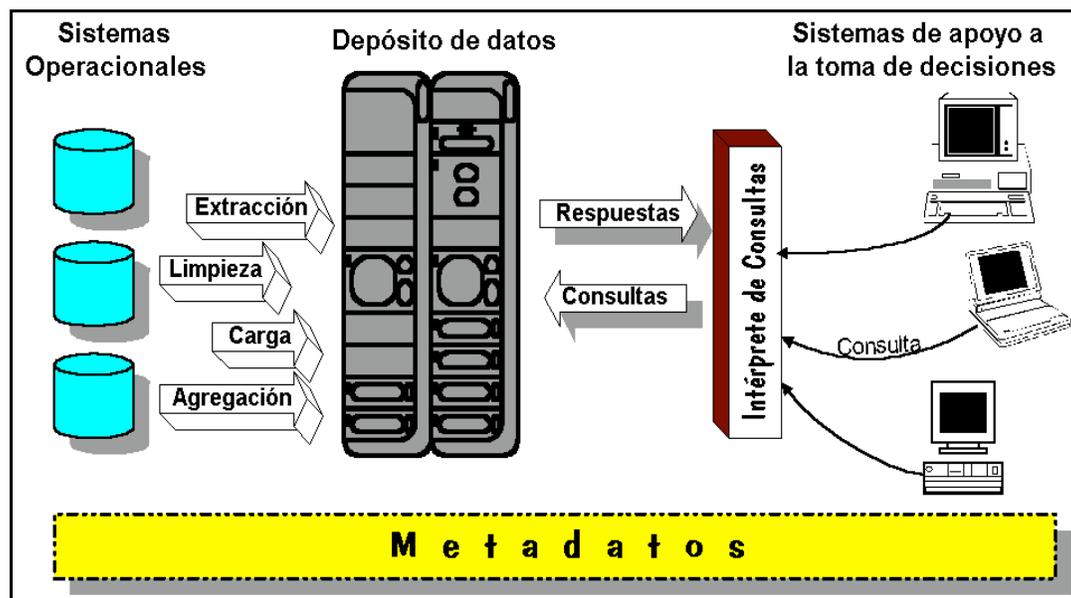


Figura No. Error! Unknown switch argument. Componentes principales en la arquitectura de un depósito

Componentes de Extracción, Limpieza y Carga

El primer paso en la fase de implementación es poblar el DW. Ello implica discernir cómo se producirá el *proceso de extracción y carga* de los datos desde los sistemas operacionales, pues en general dichos sistemas manipulan sus datos de una forma que para el DW es inconsistente, lo cual exige un análisis de *transformación y limpieza* de los datos antes de *cargarlos* en el DW.

Uno de los retos de cualquier implementación de DW está referida a dicha *transformación* de datos, siendo el objetivo del analista mezclar juiciosamente el contenido de las diversas fuentes de datos, a fin de conservar organizado el contenido del DW. Además debe:

- Integrar las fuentes de datos similares en un modelo consolidado.
- Transformar los datos provenientes de sus fuentes a una representación común.
- Seleccionar los subconjuntos de datos realmente relevantes que se requieran para el depósito - trabajando estrechamente con los analistas empresariales y ejecutivos - ya que no todos los datos provenientes de las fuentes alimentan el DW.

- Depurar los nombres de los elementos de datos para unificar la denominación, incluyendo la identificación de sinónimos y homónimos.

Esta transformación de los datos permite su enriquecimiento, ya que por medio de la combinación de diferentes atributos de diversos sistemas fuentes se crean nuevas entidades que son creativamente analizadas dentro del DW.

El proceso de extracción, transformación y carga de los datos del DW es la parte de un proyecto de DW que más recursos y tiempo consume. Por ello, al poblar el DW, uno de los métodos que pueden ayudar es el *Middleware*, que es una combinación de:

- Software de extracción (carga los datos de los sistemas OLTP).
- Proceso de transferencia de datos (envía datos, algunas veces vía FTP, a la plataforma donde el software de transformación reside).
- Software para transformación y limpieza (reconocimiento de los archivos que han sido transferidos de modo tal que estén listos para ser cargados en el DW).
- Todo proceso que añada o reemplace datos en el depósito.

Un buen diseño DW incorpora estos datos capturados de los sistemas operacionales y los utiliza para controlar el uso, seguridad, integridad y costo del DW.

Componente de Agregación

Los resúmenes o agregaciones son un componente clave en cualquier DW, pues los requerimientos de los DSS están orientados a tendencias y agregaciones más que a transacciones individuales; en tal sentido, el DW debe ofrecer niveles de resúmenes apropiados, en donde los datos sean pre-agregados y almacenados según su frecuencia de acceso.

Típicamente, los datos se cargan en el DW desde los sistemas operacionales a un nivel atómico de detalle; sin embargo, muchas de las consultas de apoyo a la toma de decisiones requieren niveles agregados y consolidados de información y no necesitan del detalle, de hecho, tratan de evitar el nivel detallado, que implicaría la agregación y recuperación de numerosos registros necesarios para la consolidación de la información, provocando tiempos de respuesta muy lentos.

Cada resumen debe describir el proceso de derivación, el detalle de las entidades, los atributos y la cantidad de cálculos ejecutados. El diseñador del DW puede evaluar el costo beneficio de ciertos resúmenes, pues en algunos casos tendrá sentido reducir costos a partir de resumir sólo aquellos valores estratégicamente interesantes para la organización; en otras palabras, que todos los valores derivados de una determinada entidad deben evaluar los resúmenes a partir de combinaciones de atributos.

Sin embargo, es literalmente imposible pre-consolidar los datos, si antes no se conoce cómo la comunidad empresarial planifica el uso de la información, resultando de vital importancia resolver acertadamente estas preguntas: *¿Cuál es la probabilidad de que un determinado resumen con gran volumen de datos nunca sea analizado?, ¿En qué forma se alteran las decisiones sobre los resúmenes si se producen cambios en las necesidades de los sistemas de apoyo a la toma de decisiones?*

Metadatos

Antes que el DW pueda ser accedido de forma eficiente, es necesario conocer qué datos están disponibles en el DW y dónde están localizados, lo cual podrán comprender los usuarios finales en la forma de metadatos, que según [Devlin 1997] son:

“Datos que describen el significado y estructura de los datos empresariales, además describen cómo son creados, accedidos y utilizados”.

En una arquitectura de distribución de datos estructurada y bien definida, las aplicaciones se unen al DW vía metadatos para apoyar a los usuarios en el uso y comprensión de los datos empresariales sin tener que conocer dónde residen los datos o en qué forma están almacenados.

La creación de los metadatos constituye un proceso costoso en el tiempo, siendo éste uno de los elementos más importantes en la arquitectura de un DW, pues describen el contenido de los datos, las definiciones de los elementos de datos, las fuentes de donde son extraídos y cómo los datos son integrados y transformados antes de ser almacenados en el depósito.

Cada implementación usa técnicas y métodos específicos para la creación de los metadatos, en dependencia de los requerimientos de cada organización, sus capacidades disponibles y los requerimientos de interfaces usuarias. Al no existir estándares para los metadatos, éstos deben ser definidos según el software seleccionado para implementar el DW y según los esquemas y subesquemas de las BD de los sistemas transaccionales.

Los metadatos son para el DW lo que el catálogo (por materia, autor, etc.) para una biblioteca, es decir, permiten identificar el significado y localización de los datos dentro del depósito. Entre sus funciones están almacenar:

- La extracción y transformación de los datos⁸, con su seguimiento histórico.
- La ubicación y descripción de servidores y bases de datos.
- Los algoritmos de consolidación y modelaje utilizados.
- Las estadísticas de utilización de los datos.
- La estructura y contenido del DW.
- Los niveles de resúmenes, el método utilizado y sus tablas de registros.
- Los registros de los cambios a través del tiempo⁹.
- Las reglas empresariales¹⁰.
- Las reglas de control de acceso y seguridad.
- Las jerarquías de atributos, etc.

Intérprete de consultas

Es el componente por el cual todas las consultas al depósito deben pasar, ya que entiende los metadatos que definen la estructura utilizada en el DW. Por lo general, es utilizado para esconder - de la vista de los usuarios - la complejidad del DW. Normalmente, reside en un servidor e

⁸ Especificaciones de las transformaciones hechas a los datos desde sus fuentes de origen.

⁹ Registro de cuándo nuevos elementos de datos han sido añadidos al DW y cuán históricos son los elementos de datos que han sido eliminados o resumidos.

¹⁰ Lineamientos formales que establecen las instituciones para resolver e implementar sus políticas.

intercepta las consultas de los analistas de datos, expertos empresariales o gerentes encargados de tomar las decisiones, las modifica y las pasa a la BD del DW.

Para resumir, sus funciones primarias son:

- Esconder a los usuarios la complejidad de los datos.
- Interceptar las consultas y direccionarlas.
- Modificar las consultas para optimizar el rendimiento.

Sistemas de apoyo a la toma de decisiones

No existe una herramienta única capaz de satisfacer todos los requerimientos, por tal razón, se acostumbra integrar un conjunto de herramientas y facilidades dependiendo del tipo de usuario y de sus requerimientos específicos, éstas pueden incluir generadores de reportes, software de consultas, OLAP y herramientas de minería de datos.

En general, los DSS son una combinación de herramientas, paquetes de software y programas computacionales personalizados que apoyan el análisis estratégico de la información; para lo cual requieren de los siguientes aspectos:

- Seguridad (tanto a nivel de tipo de dato como de valor).
- Facilidad en la construcción de consultas en pantallas desde entidades, atributos y agregaciones del DW.
- Facilidad de transparencia para acceder a datos locales, centrales y a nivel de escritorio.
- Modos de operación para expertos y novatos.
- Importa las capacidades de un DW a herramientas de escritorio.

2.3 Arquitectura por niveles

En un reciente estudio de Gartner Group [Barquin 1997] el 73% de los gerentes consultados dijeron que no podían acceder fácilmente a los datos de sus propias bases de datos corporativas, lo cual constituía un inconveniente para satisfacer sus objetivos empresariales.

Una solución comúnmente empleada al querer mejorar el acceso a los datos preservando las inversiones en los sistemas centralizados (*minis o mainframes*), es aumentar los sistemas legados actuales con paquetes de aplicación que se ejecutan muy eficientemente en servidores. Adicionalmente a los *mainframes*, se utilizan aplicaciones cliente/servidor que permiten el acceso transparente a los datos e implementan nuevos servicios que aportan ciertos beneficios.

2.1.3.1 Arquitectura de dos niveles

Los DW apoyan una estrategia que preserva la inversión en los sistemas legados debido a que en muchos casos las arquitecturas de los DW mantienen y utilizan la infraestructura existente (típicamente centralizada).

En esta opción arquitectónica para construir un DW se utiliza un equipo centralizado como servidor (ver Figura No. 3). Su definición de dos niveles se atribuye a que utiliza dos componentes: *front end* (cliente) y *back end* (servidor).

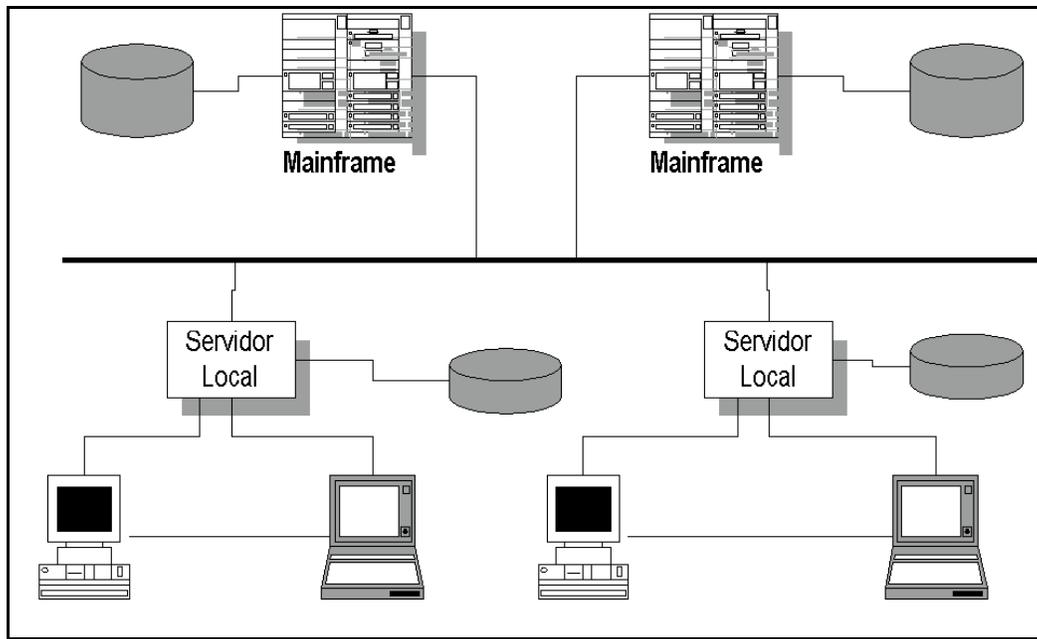


Figura No. **Error! Unknown switch argument.** Arquitectura de dos niveles

La arquitectura ilustrada basada en una arquitectura de dos niveles, resulta atractiva debido a que utiliza el legado existente como servidores de bases de datos y requiere una mínima inversión en hardware y software. Una forma de tratar con esta situación es no tocar el legado de sistemas heredado y construir sistemas en el componente *front end* que suministren fácil acceso a los datos que se encuentran en el *mainframe*.

El componente de *front end* y la red de área local (LAN) se basan en aplicaciones que se construyen a partir de herramientas como [Barquin 1997]:

- Visual Basic de Microsoft.
- Power Builder de Sybase/Powersoft.
- Delphi de Borland.
- SQLWindows de Gupta, etc.

La función de estas aplicaciones *front end* es esconder la complejidad de los sistemas basados en *minis* y *mainframes* existentes. Las aplicaciones basadas en *mainframes* que han sido utilizadas por muchos años, algunas veces poseen altas tasas de redundancia y gradualmente acumulan múltiples sistemas computacionales, cada uno de los cuales pudiera tener definiciones de datos incompatibles y redundantes. La redundancia y la escasez de definiciones consistentes dificultan el acceso usuario al legado de los datos de los sistemas existentes y pueden impedir el desarrollo de nuevos servicios de *front end*.

Las aplicaciones *front end* construidas con herramientas cliente/servidor como Visual Basic, Delphi o SQLWindows brindan una interfaz gráfica y amigable que satisface funciones específicas como: procesamiento de órdenes de compras, servicio al cliente u otros. Las aplicaciones *front end* proveen acceso transparente al legado de los sistemas existentes y esconde a los usuarios finales, la complejidad y la escasez de consistencia que puede presentarse en las bases de datos de los equipos centralizados.

En la medida en que el número de usuarios se incrementa, los requerimientos de accesos a los datos de las aplicaciones *front end* imponen una creciente carga para los *mainframes*, por lo que para satisfacer un adecuado tiempo de respuesta, debe incrementarse el procesamiento y la capacidad de acceso a los datos de los *mainframes*.

Esta arquitectura no es escalable por estar limitada a la capacidad de los *mainframes*; no obstante, fomenta el desarrollo de “clientes gordos”¹¹ en los sistemas *front end*.

2.1.3.2 Arquitectura de tres niveles

Una alternativa arquitectónica es utilizar la estructura de información multinivel de la organización, tal como se muestra en la Figura No. 4. Esta arquitectura flexible soporta un amplio rango de servicios integrados, en el cual la interfaz usuaria, los procesamientos de funciones empresariales y las funciones de administración de las bases de datos son particionados en procesos separados que son distribuidos a lo largo de la estructura informacional de la organización.

La arquitectura de tres niveles es una de las más utilizadas para la construcción de un DW. Las fuentes de datos en los equipos centralizados residen en el tercer nivel (parte superior del gráfico). Las reglas empresariales que son compartidas a lo largo de toda la organización tales como la base de datos para el DW, son almacenadas en servidores de alta velocidad en el nivel 2. Las interfaces gráficas orientadas a usuarios finales corren en LAN basadas en PCs o estaciones de trabajo en el nivel 1.

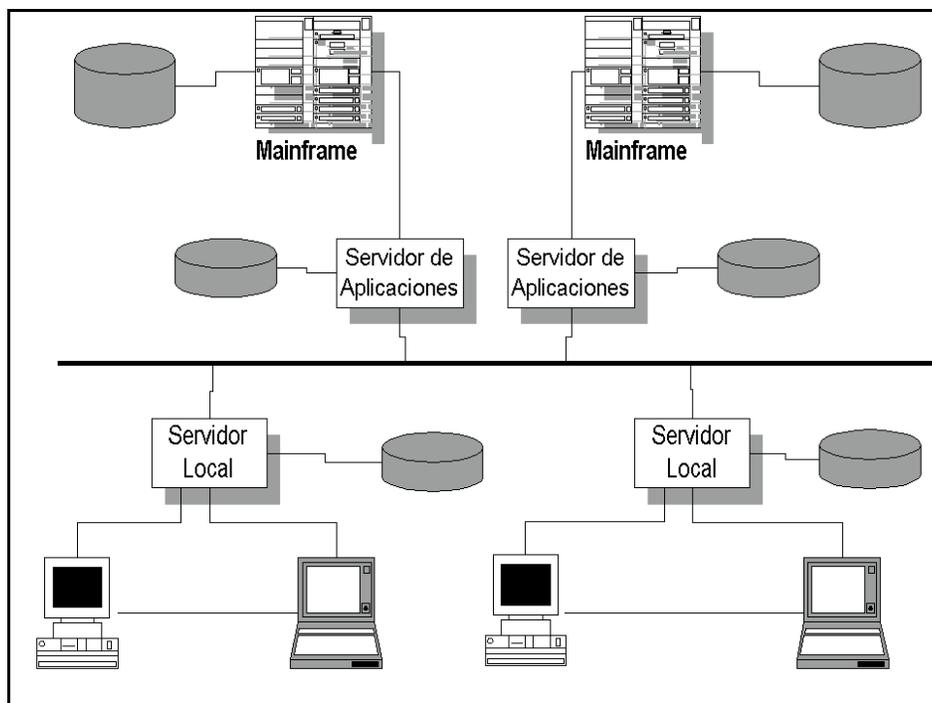


Figura No. **Error! Unknown switch argument.** Arquitectura de tres niveles

¹¹ Son aquellos que hacen un procesamiento prominente de aplicaciones desde sus estaciones de trabajo.

Las bases de datos que utilizan los DW pueden ser del tipo relacional, tales como: Oracle, Sybase, Informix, DB2.

Las aplicaciones que requieren análisis de grandes cantidades de datos multidimensionales pudieran utilizar herramientas OLAP especializadas, tales como [Barquin 1997]:

- Essbase de Arbor Software Corp.
- Express de Oracle.
- OmniWarehouse de Praxis International.
- Red Brick de Red Brick System.

Las herramientas de usuario final son utilizadas para apoyar el acceso a altas velocidades y el análisis de los datos en las bases de datos. Las técnicas de acceso empleadas pueden incluir [Barquin 1997]:

Consultas relacionales como:

- SQLWindows de Gupta.
- Esperant de Software AG.
- Power Object de Oracle.

Vistas multidimensionales de datos relacionales como:

- SAS/EIS del Instituto SAS.
- MetaCube de Stanford Technology Group (división de Informix Corp.)

Herramientas de propósito especial para el proceso de toma de decisiones como:

- Express de Oracle.
- Lighthouse de Pilot Software.

Llegados a este punto es importante reconocer que no hay una solución correcta única en cuanto a lo que se refiere a la arquitectura de un DW. En la construcción del depósito para algunas organizaciones, la definición a partir de dos niveles constituye una atractiva solución debido a sus reducidos costos y complejidad. Para otras que requieren altos rendimientos y escalabilidad, la arquitectura de tres niveles pudiera ser la apropiada.

2.4 Arquitectura por bloques

La arquitectura expuesta por los autores - en [Harjinder 1996] - se describe desde un punto de vista sencillo y abstracto del siguiente modo:

- Un conjunto de datos extraídos de bases de datos operacionales.
- Un software que prepara los datos para que los accedan los usuarios.
- Un conjunto de aplicaciones y herramientas que ejecutan consultas y análisis complejos.

A criterio de sus autores esta arquitectura divide los componentes del DW en: bloques y capas (ver Figura No. 5), lo cual permite:

- Construir el DW, el mercado de datos o ambos.
- Seleccionar pocas fuentes de información para construir el DW, reservándose el derecho de admitir más en el futuro.

- Definir una plataforma múltiple. La arquitectura al estar conformada en varios niveles, puede usar distintos tipos de plataformas para los componentes arquitectónicos. Por ejemplo, las fuentes de datos y el depósito pueden basarse en macrocomputadoras, los mercados de datos en una plataforma cliente/servidor, y los accesos y usos por parte de los usuarios finales en estaciones de trabajo.
- Seleccionar los componentes tecnológicos de la capa de infraestructura y transporte.

Los bloques están constituidos por las fuentes de datos, la construcción del depósito (DW) y la construcción del mercado de datos (MD), así como por el bloque de acceso y uso del DW. Las capas están constituidas por la administración de datos y de metadatos, así como por el transporte y la infraestructura.

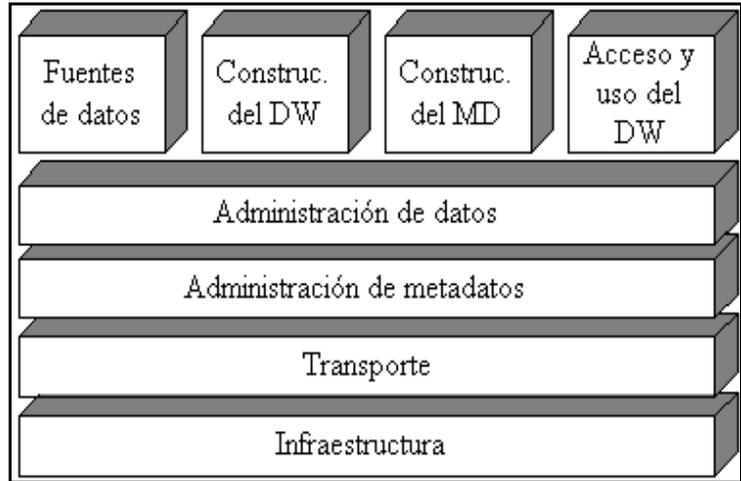


Figura No. **Error! Unknown switch argument.** Arquitectura por bloques

Los bloques de construcción del DW y del MD, acceso y uso del depósito, junto con las capas de administración de datos y metadatos, representan las áreas de nueva inversión, mientras que las fuentes de datos e infraestructura representan las inversiones existentes dentro de la empresa.

Componentes por bloques

Bloque de fuentes de datos

Este bloque se divide en las categorías:

- Datos de producción. Se refieren a las bases de datos operacionales que contienen la información recopilada de las aplicaciones operacionales.
- Datos de herencia. Son datos que no son necesarios para apoyar aplicaciones operacionales actuales, pero tienen un gran valor histórico para el análisis de tendencias y son útiles para la minería de datos.
- Sistemas internos de oficina. Son fuentes de datos para sustentar el análisis en los departamentos que no están almacenadas en una base de datos operacional, tal como las formas no electrónicas, documentos de procesadores de palabras, reportes, etc.
- Fuentes externas. Son aquellos datos que la empresa no posee ni controla. Pueden tener fuentes electrónicas (resúmenes de análisis competitivo del mercado de firmas de investigación) y no electrónicas (los reportes de consultores, los artículos y revistas).
- Metadatos para fuentes. Incluye el nombre del dato, definición del contenido, ubicación dentro del depósito, parámetros de control, etc.

Bloque de construcción del DW

Como muestra la Figura No. 6, este bloque consta de los siguientes componentes:

- El componente de *refinamiento* es responsable de estandarizar y limpiar la información extraída de las fuentes de datos seleccionadas.
- El componente de *reingeniería* es responsable de preparar los datos para que sean congruentes con las necesidades de análisis empresarial. Esta reingeniería implica:
 - La integración de diferentes tipos de datos de sistemas múltiples.
 - La traducción y formato de datos de diferentes fuentes.
 - La transformación y reubicación de los datos almacenados.
- El componente de *depósito de datos* es el encargado de modelar, validar y resumir las consultas provenientes de los usuarios.

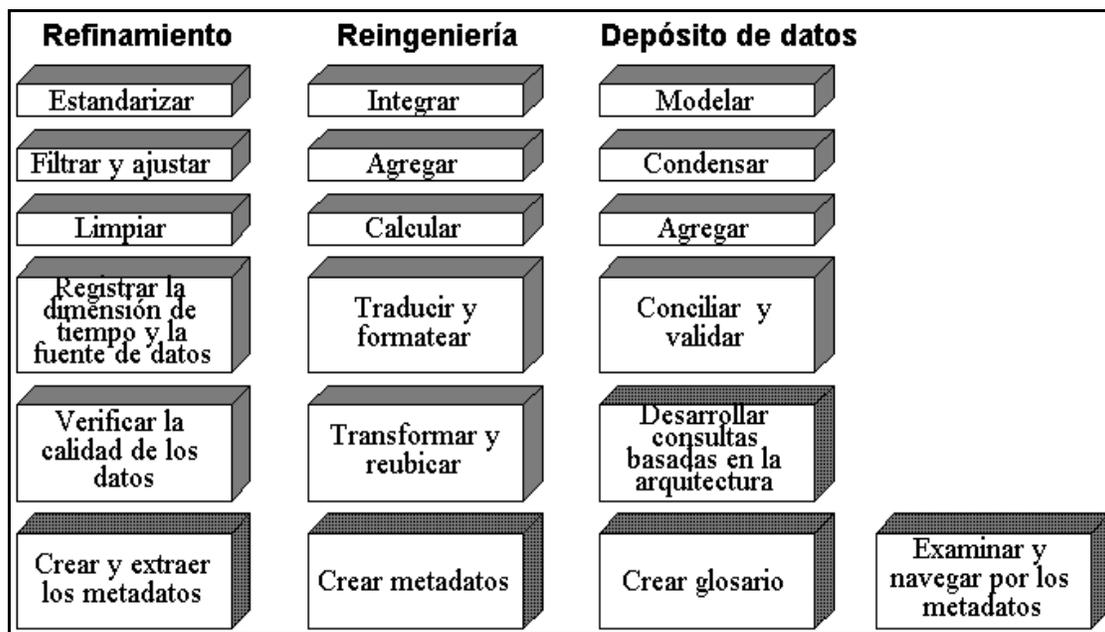


Figura No. Error! Unknown switch argument. Bloque de construcción del DW

Bloque de construcción del mercado de datos

Este bloque se utiliza para crear un mercado de datos a partir del contenido del DW, el cual requiere de los siguientes componentes (ver figura No. 7):

El componente de *refinamiento y reingeniería* es responsable de:

- El ajuste de la información extraída del DW.
- La integración del DW en las áreas empresariales nuevas o redefinidas.
- La creación de nuevos resúmenes o adiciones.

El componente de *creación del mercado de datos* posee los mismos componentes de los bloques de creación del DW descrito antes, su diferencia radica en el enfoque de cada uno de los componentes. Por ejemplo, el bloque de modelado en el mercado de datos se concentra en las necesidades departamentales.

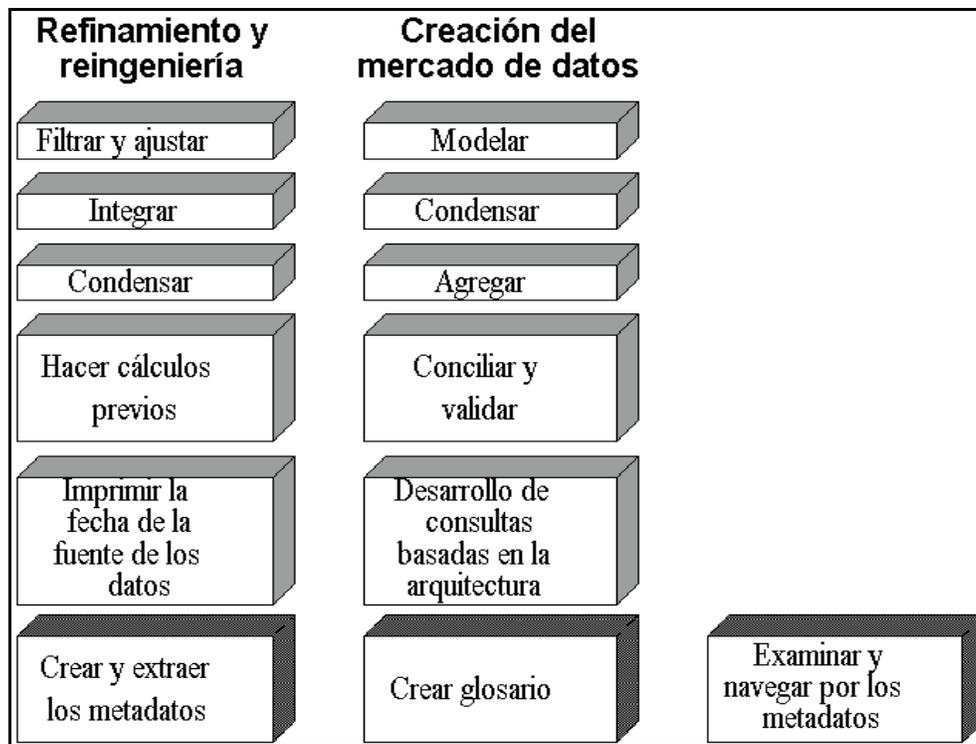


Figura No. **Error! Unknown switch argument.** Bloque de construcción del mercado de datos

Bloque de acceso y uso del DW

La figura No. 8 muestra los componentes en detalle:

El componente de *acceso y recuperación* ayuda al usuario empresarial a:

- Entender qué es el DW (nombre, descripción, contenido, valor, fuente, etc.).
- Entender el contexto de los datos y la transformación que sufren dentro del DW.
- Determinar la ubicación de los datos.
- Verificar la confiabilidad de los datos.

El componente de *análisis y reporte* es responsable de la familia de herramientas y aplicaciones necesarias para explotar el DW, y se clasifican como se presenta en la figura No. 8.

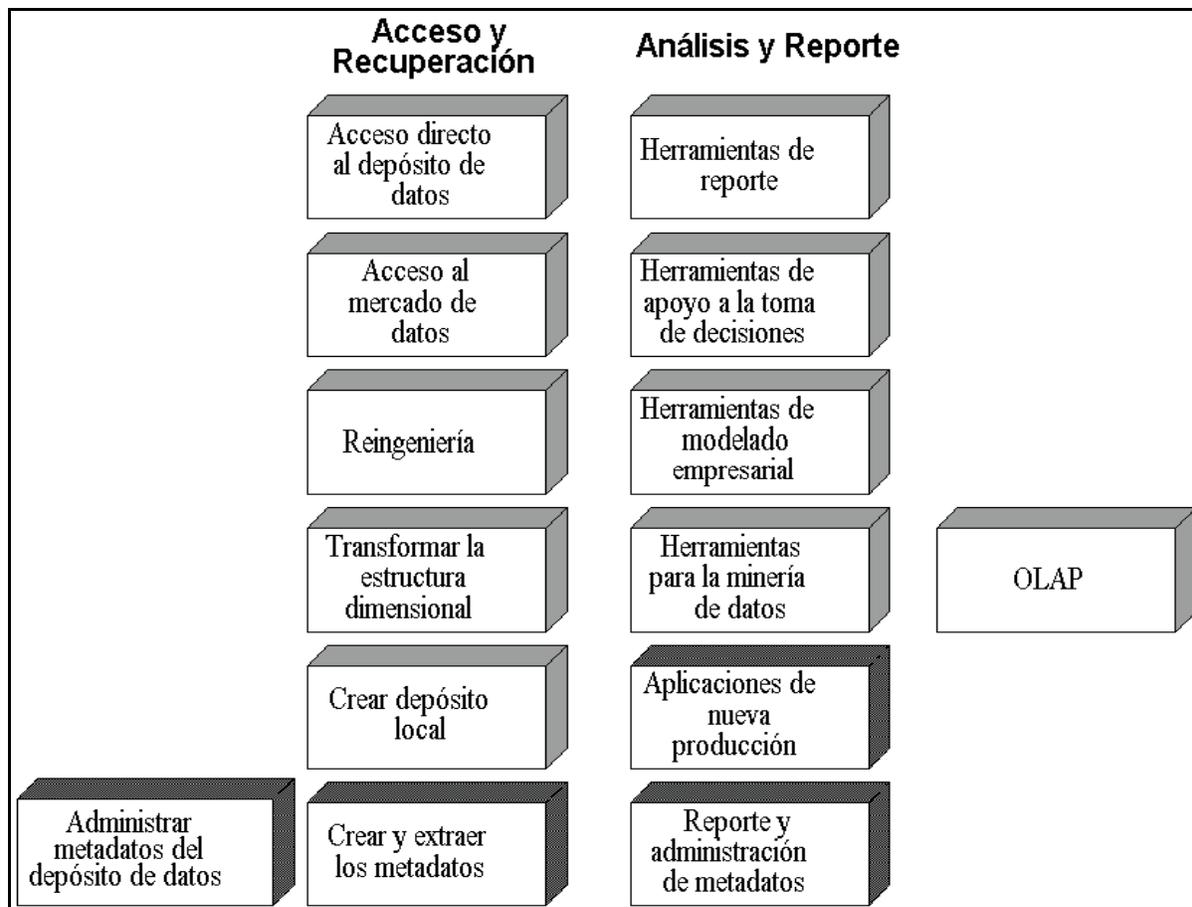


Figura No. **Error! Unknown switch argument.** Bloque de acceso y uso del DW

Componentes por capas

Tal como se muestra en la figura No. 5, las capas de administración de datos y de metadatos, representan el ambiente necesario para implementar los bloques y se corresponden con las actividades relacionadas con la extracción y carga del DW. Las otras capas corresponden a servicios esenciales para los bloques del DW como son la infraestructura y el transporte.

Capa de administración de datos

En esta capa los elementos de interés son:

- Extraer los datos apropiados, para su posterior refinamiento e incorporación al DW.
- Captar los datos que han sido cambiados en las fuentes de datos operacionales para después llevarlos al DW.
- Verificar las fuentes de datos nuevos o más actualizados.

El desempeño de esta capa tiene lugar a través de los siguientes elementos:

- Extracción y administración de solicitudes de consulta
- Carga, almacenamiento y actualización de sistemas
- Seguridad y autorización de sistemas
- Archivo, restauración y purga de sistemas

Capa de administración de metadatos

Esta capa es responsable de:

- Definir los estándares de los datos.
- Registrar los metadatos capturados y creados en los bloques de refinamiento y reingeniería.
- Describir las consultas y reportes diseñados.
- Describir las reglas para preparar y programar el ciclo de actualización y duplicación.

El desempeño de esta capa tiene lugar a través de los siguientes elementos:

- Administración del DW, del esquema del mercado de datos y del glosario
- Administración de la extracción, creación, depósito y actualización de los metadatos
- Administración de las consultas predefinidas, los reportes y los índices
- Administración de la actualización y la duplicación
- Administración de las conexiones, generación de archivos, restauración y purga

Capa de transporte

La capa de transporte soluciona los puentes de comunicación necesarios entre los distintos bloques de la arquitectura de referencia, esta capa utiliza para ejecutar su función los siguientes componentes:

- Red para transferencia y entrega de datos
- Herramientas de *middleware* y sistemas cliente/servidor
- Sistemas de duplicación
- Sistemas de seguridad y autenticación

Capa de infraestructura

Esta capa está compuesta por los siguientes elementos:

- Administración de sistemas
- Administración del flujo de trabajo
- Sistemas de almacenamiento
- Sistemas de procesamiento

2.5 Arquitectura basada en el modelo empresarial

La arquitectura de un DW requiere la utilización de un conjunto de técnicas diferentes a las utilizadas en aplicaciones tradicionales u operacionales, debido a que [Devlin 1997]:

- El alcance del DW eventualmente abarca a toda la organización.
- El DW contiene el registro histórico de la organización.
- La fuente de los datos está dispersa, y difiere en su estructura y contenido.

Desde el punto de vista empresarial, esta arquitectura basada en los datos empresariales, suministra la forma en que son utilizados los datos y facilita el diseño de sistemas computacionales que apoyan el funcionamiento de la empresa, lo que proporciona:

- Un registro preciso del significado de los datos empresariales.
- Una estructura de datos que permite administrar la organización.

La información contenida en un modelo de este tipo puede ser usada por los diferentes niveles empresariales. Al más alto nivel el modelo permite a los gerentes planificar las estrategias empresariales y estimar su impacto; por otra parte, es importante en el entorno operacional para el diseño y el desarrollo de procesos.

Estructura de un modelo de datos empresarial

Es de notar en la Figura No. 9, que la estructura típica de un modelo de datos empresarial (EDM) tiene una forma piramidal dividida en niveles; en el más alto nivel la cantidad de información es mínima, la cual se incrementa a medida que decrece en nivel de modelado. El EDM representa todos los datos necesarios para la organización desde el más alto nivel empresarial hasta el diseño lógico y físico de los datos.

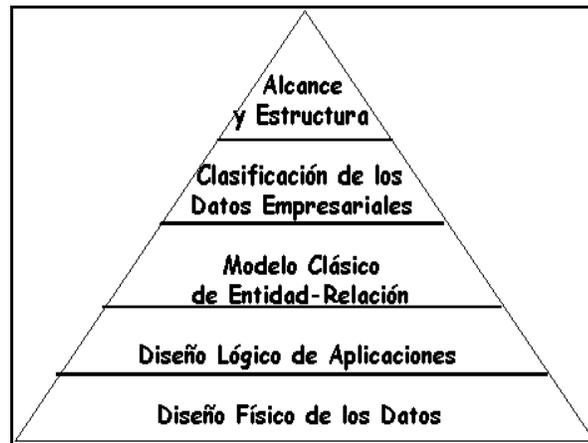


Figura No. Error! Unknown switch argument. La estructura de un EDM

Niveles del modelo de datos empresarial

➤ *Alcance y Estructura*

Representa el punto de vista más alto y consolidado de los datos empresariales, generalmente identifica de 10 a 20 conceptos que representan los temas y áreas sustantivas para la empresa. Estos conceptos son definidos como entidades y expresados en una terminología empresarial que es comprendida y utilizada por toda la organización.

➤ *Clasificaciones de los datos empresariales*

Este nivel constituye un medio para definir el contenido de los diferentes conceptos y categorizaciones de acuerdo con las reglas empresariales, permitiendo a las diversas partes de la organización revisar sus conceptos con un mayor nivel de detalle.

Un ejemplo de categorías es ‘producto’ y ‘partes involucradas’, donde el concepto ‘producto’ se explica por sí solo, mientras que ‘partes involucradas’ se refiere a personas u organizaciones con la cual la empresa posee algún tipo de relación legal.

➤ *Modelo Clásico de Entidad-Relación*

Las clasificaciones de los datos empresariales sirven para generar el modelo clásico de entidad-relación, siendo éste un modelo completo que identifica y describe todas las entidades, atributos y relaciones usadas en todo el entorno empresarial.

➤ *Diseño lógico de aplicaciones*

Está en estrecha relación con el modelo clásico de entidad-relación, su diferencia estriba en que se particiona en diferentes vistas que constituyen las bases de aplicaciones específicas.

Con el objetivo de satisfacer las necesidades de una aplicación empresarial, una misma entidad podría aparecer varias veces en el diseño lógico de una determinada aplicación.

➤ *Diseño físico de los datos*

En este nivel se aplican las restricciones de implementación física, tales como la distribución física de los datos entre localizaciones, las fuentes de los datos, el rendimiento, etc. Estas restricciones deben ejecutarse independientemente de las consideraciones de utilización empresariales, con el propósito de garantizar que cualquier cambio tecnológico no provoque un impacto sobre el modelo lógico.

Un DW debe considerar explícitamente los aspectos temporales de los datos que contiene, proporcionando por definición una vista histórica de los datos empresariales; debido a que los datos cambian fundamentalmente a nivel de campo (atributo), es posible representar el tiempo a este nivel o a cualquier otro nivel en la estructura, tal como registro, tabla, archivo, etc. dependiendo de la granularidad requerida.

El volumen de datos históricos por almacenar puede propiciar beneficios. De hecho, si todos los datos son almacenados a su mayor nivel de detalle y nunca son eliminados, entonces todas las posibles consultas futuras podrán ser satisfechas; pero este enfoque es difícil de justificar bajo un simple análisis de costo y beneficio.

Por lo tanto, el aspecto clave es sopesar los costos de almacenamiento y administración de datos históricos, teniendo en cuenta que según la Figura No. 10, los datos resumidos son más valiosos y utilizados en su horizonte a largo plazo en su comparación con los datos detallados, ya que los ejecutivos se centran en los datos agregados para el proceso de toma de decisiones.

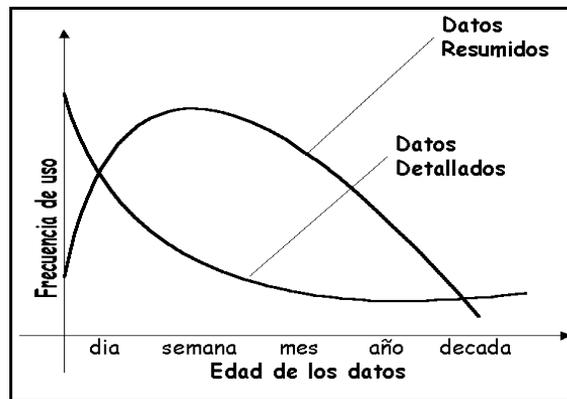


Figura No. Error! Unknown switch argument. Curva de utilización de los datos

La arquitectura lógica de un DW, enfocada desde la perspectiva de la estructura de datos puede ser dividida en [Devlin 1997]:

- Sistemas Operacionales (SO).
- DW Empresarial (BDW).
- DW Informacional (IDW).

La Figura No. 11 muestra la arquitectura y su relación con los diferentes tipos de datos.

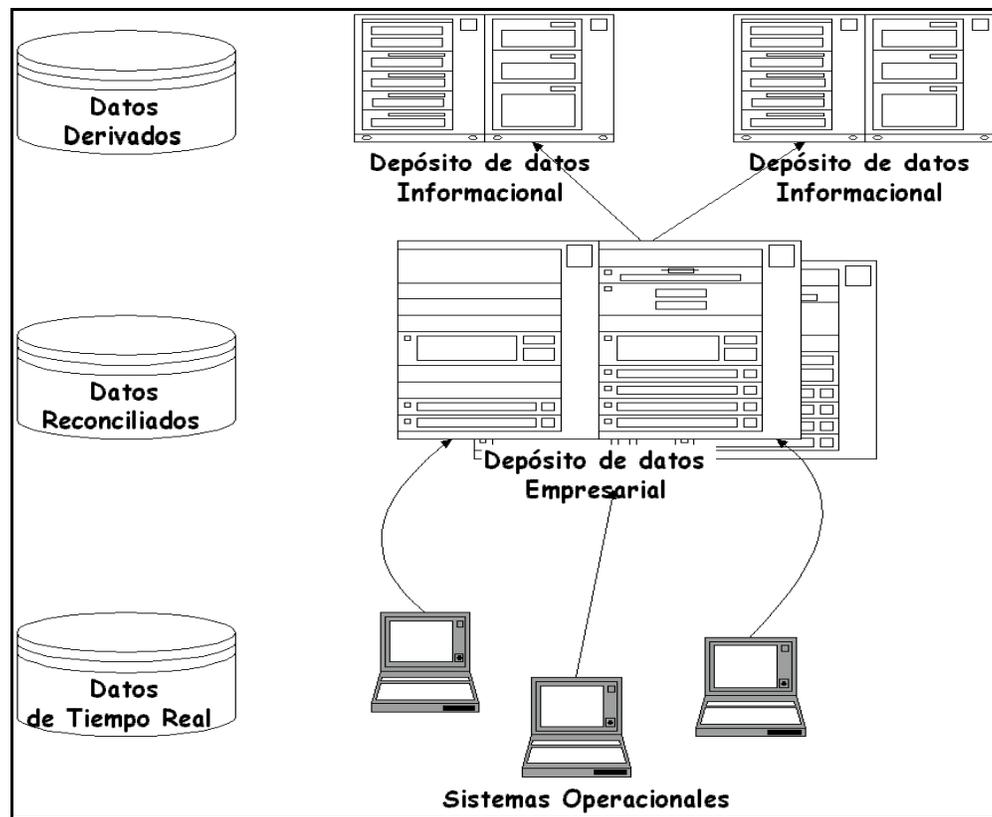


Figura No. **Error! Unknown switch argument.** Arquitectura lógica de un DW

Sistemas Operacionales (SO)

Constituyen las aplicaciones necesarias que se ejecutan para que la empresa funcione diariamente y utilice sus datos, bases y archivos de tiempo real. Estas aplicaciones existen en una diversidad de formatos y localizaciones, siendo tanto homogéneas como heterogéneas. Las aplicaciones más modernas son cada vez más heterogéneas y distribuidas, y están implementadas en diferentes entornos cliente/servidor.

Si bien algunos proyectos de reingeniería tratan de simplificar este entorno para hacerlo más homogéneo, hay un conjunto de factores que niegan este esfuerzo:

- La tendencia creciente hacia sistemas distribuidos favorece la diversidad tecnológica.
- El incremento de la autonomía en los departamentos y el conocimiento computacional reduce el control que el departamento de sistemas de información puede ejercer.
- La necesidad de desarrollar aplicaciones mucho más rápidas influye negativamente en la estandarización de los sistemas operacionales.

La conclusión reside en que independientemente del entorno en el que se desarrollen los sistemas operacionales, éstos constituyen la fuente de los datos que poblarán el DW y en tal sentido revisten una importancia crucial.

DW Empresarial (BDW)

Constituye una implementación centralizada de datos reconciliados¹², que son diseñados para ser el punto de control y la fuente de datos sustantiva disponible para todos los usuarios finales. Se caracteriza por ser detallada, histórica, consistente, modelada y normalizada.

El BDW se implementa en un entorno completamente relacional, constituyendo éste el mejor entorno para la naturaleza normalizada y modelada de los datos. En teoría, los BDW pueden ser distribuidos; no obstante, debido a que el proceso de reconciliación requiere integrar grandes cantidades de datos, es más natural que un proceso con tales características se adapte mejor a entornos no-distribuidos.

Los aspectos organizacionales contribuyen a realizar implementaciones centralizadas, debido a que los BDW tratan de ser el punto de control donde se asegura la calidad e integridad de los datos, antes de que estén a disposición de la comunidad de usuarios finales. Tal responsabilidad debe delegarse probablemente a un departamento centralizado de sistemas de información, ya que es más fácil de administrar si se concentra en un sólo lugar.

Otra importante consideración para el almacenamiento de la información corporativa es la seguridad de los BDW, basada en una seguridad física única o un entorno centralizado, ya que éstos poseen de forma integrada todos los datos empresariales.

DW informacionales (IDW)

Constituyen aplicaciones que los usuarios finales utilizan para administrar su organización en los períodos de tiempo y en la magnitud de datos que satisfagan sus requerimientos, las cuales incluyen técnicas de generación de informes, análisis y pronóstico de los hechos de la organización.

Este entorno es altamente distribuido y se encuentra bajo implementaciones de arquitecturas cliente/servidor. Los IDW poseen datos derivados¹³, los cuales se diseñan para satisfacer las necesidades de los usuarios finales: individualmente, por grupos, departamentos o divisiones. Puede contener datos tanto a nivel detallado como agregado, con una estructura optimizada para responder a consultas en líneas, ya sean 'ad-hoc' o predefinidas.

A modo de conclusión, las organizaciones en la selección y diseño de la arquitectura para el DW, deben investigar un amplio rango de alternativas arquitectónicas y seleccionar aquella que satisfaga sus requerimientos empresariales específicos:

- Para algunas organizaciones la definición de dos niveles es una solución atractiva debido a que minimiza los costos y la complejidad en su construcción.
- Para otras organizaciones que requieran grandes rendimientos y escalabilidad, la arquitectura de tres niveles pudiera ser la más apropiada.
- Por su parte, aquellas organizaciones que poseen requerimientos más modestos, como son la necesidad de DW para departamentos o secciones individuales, su mejor selección sería construir mercados de datos.

¹² Datos transformados e integrados provenientes de los sistemas operacionales.

¹³ Datos utilizados para analizar y administrar la empresa, por ejemplo, el análisis costo beneficio trimestral de los últimos 12 trimestres.

Capítulo 3

Proceso de implementación de un depósito de datos

“El éxito no se alcanza por aquellos que se adaptan al cambio, sino por los que sean capaces de crecer en dicho cambio”.

El objetivo primordial de un DW es satisfacer las necesidades de información. Sin embargo, su proceso de implementación tiene un gran reto que es cómo convertir la descripción que hace la administración de sus problemas u objetivos, en una lista de consultas y resultados que se puedan extraer del DW para resolver problemas empresariales específicos.

Con este fin y tomando como base la arquitectura genérica expuesta en el capítulo anterior, se presenta a continuación un proceso de implementación de un DW, el cual debe examinar los siguientes factores críticos de éxito:

Comprometer a la máxima gerencia empresarial

El nivel más alto de dirección empresarial debe estar consciente de los objetivos, costos, beneficios y riesgos del proyecto, de lo contrario sería muy difícil garantizar los recursos necesarios (especialmente financieros y humanos) para el desarrollo exitoso del proyecto. Una buena táctica para garantizar este objetivo es que un miembro de la máxima gerencia pertenezca al equipo de trabajo encargado del proyecto.

Adecuada justificación financiera

Los proyectos de DW en sus inicios son muy costosos tanto por el equipamiento como por el adiestramiento del equipo de trabajo. La táctica básica es realizar un buen análisis económico y pronósticos financieros que permitan aprobar el presupuesto suficiente para satisfacer los recursos requeridos por el proyecto, así como determinar el período de retorno de la inversión.

Designar un director de proyecto a tiempo completo

La persona que se responsabilizará con el desarrollo del proyecto no tiene necesariamente que ser un especialista tecnológico, pero sí se exige de él la capacidad y la experiencia en la administración de proyectos informáticos.

Apropiado equipo de trabajo

Comprometer con el proyecto a los especialistas en temas empresariales que conforman la base para lograr la satisfacción de los objetivos empresariales, así como de especialistas del área de sistemas de información con experiencia en interfaces, bases de datos, redes, etc.

Herramientas de trabajo productivas

Es esencial la utilización de herramientas de software tales como una herramienta CASE que permitan administrar el modelo requerido, una herramienta de administración de proyectos que facilite controlar las tareas y fechas de cumplimiento acordadas, etc.

Existe una marcada diferencia entre administrar un proyecto informático tradicional y administrar un proyecto de DW, ya que muchas prácticas y reglas tradicionales de la TI son desechadas para la construcción de un DW; por consiguiente, la mayoría de las organizaciones requieren de una curva de aprendizaje pronunciada para adaptar esta nueva forma de pensamiento.

Con el fin de estructurar los pasos necesarios para construir un DW, se propone un proceso de implementación subdividido en etapas y pasos. Las etapas del proceso de implementación son las siguientes:

Etapas 1 Fundamentación.

Etapas 2 Modelaje.

Etapas 3 Preparación e Implementación.

Etapas 4 Operación.

3.1 Etapa 1 Fundamentación

La primera tarea del equipo de trabajo seleccionado es evaluar el posible impacto que el DW tendrá en la actividad empresarial y demostrar por qué está fundamentado un proyecto de tal magnitud, considerando los siguientes elementos:

- Establecer correctamente los objetivos, alcance y resultados esperados del proyecto.
- Estimar y justificar los costos del proyecto de DW.
- Establecer los recursos necesarios para iniciar y ejecutar las actividades requeridas.
- Aprobar el financiamiento y presupuesto a fin de acometer el proyecto.

En esta etapa debe tomarse la decisión de proceder o no con el proyecto, especificando los beneficios que la empresa espera obtener tanto a corto plazo como a largo plazo, teniendo en cuenta que los objetivos primarios se centrarán en las metas empresariales.

Una vez que el proyecto es aprobado deben ejecutarse los siguientes pasos:

- 3.1.1 Definir las fronteras del sistema.
- 3.1.2 Identificar los objetivos empresariales del DW.
- 3.1.3 Identificar las consultas potenciales.
- 3.1.4 Identificar las fuentes de datos potenciales.
- 3.1.5 Pronosticar los volúmenes iniciales.
- 3.1.6 Analizar costo beneficio.
- 3.1.7 Obtener la aprobación del presupuesto.
- 3.1.8 Desarrollar un plan de contingencias.
- 3.1.9 Suministrar capacitación.

3.1.1 Definir las fronteras del sistema

La definición del alcance del sistema está limitada por el costo estimado total del proyecto. En su definición se deben especificar los aspectos que repercuten de manera directa en toda la organización, evitando los detalles innecesarios.

Para definir el alcance del área bajo estudio y las entidades externas con las cuales ésta debe interactuar, se utiliza un diagrama de contexto como forma de representación inicial de los flujos de la empresa solamente como entradas y salidas.

3.1.2 Identificar los objetivos empresariales del DW

El objetivo principal de esta tarea es identificar y esclarecer los objetivos y resultados esperados con el proyecto de DW, planeando sesiones de trabajo en grupo y entrevistas estructuradas con la comunidad empresarial, con el propósito de que la recolección de la información acerca de los objetivos empresariales sea cuantificable y medible, para que dichos objetivos empresariales sean satisfechos por el DW.

El núcleo fundamental de esta tarea es elaborar un informe de objetivos que identifique el propósito del proyecto por realizar, como parte de la administración del proyecto.

3.1.3 Identificar las consultas potenciales

No es posible conocer a priori todas las consultas que deberá satisfacer el DW, pero el tratar de identificar aquellas que ocurren con mayor frecuencia, será algo que facilitará el buen funcionamiento del DW.

Con el propósito de identificar los diversos tipos de requerimientos y restricciones de todos los niveles de la organización, a los cuales el DW va a tener que dar servicio, es recomendable elaborar un informe de consultas. Este informe reviste una gran importancia para la tarea de diseño del esquema del DW, así como para la identificación de las fuentes de datos potenciales.

3.1.4 Identificar las fuentes de datos potenciales

Esta tarea busca acercarse lo más posible a las fuentes donde se generan los datos, con el objetivo de identificar los datos necesarios que permitan la carga y explotación del DW, para lo cual se elabora un listado de las fuentes de datos.

3.1.5 Pronosticar los volúmenes iniciales

Esta tarea consiste en realizar un estimado de los volúmenes y tasas de crecimiento asociadas con los principales elementos de la organización (no de los sistemas). Para obtener estos estimados son de gran ayuda preguntas como las siguientes:

- Cuántas propiedades diferentes (oficinas, sucursales, almacenes) posee la organización.
- Cuántos productos elaboran.
- Cuántos pedidos son procesados por mes, semanas y días.
- Cuál es el número promedio de elementos por pedidos.

Los resultados son utilizados para estimar los requerimientos de almacenamiento para la base de datos que utilizará el DW. Esta información puede ser obtenida de entrevistas realizadas a los expertos empresariales o inferida del análisis de los sistemas fuentes. La información de los expertos puede diferir de las inferencias que se hagan a partir de las fuentes potenciales; no obstante, la existencia de tales diferencias ayuda a una mejor comprensión del tamaño estimado para el DW.

3.1.6 Analizar costo beneficio

Su objetivo fundamental es determinar si el proyecto es factible económicamente y justificable desde el punto de vista financiero. El análisis costo beneficio permite determinar si los beneficios esperados por la ejecución del proyecto son superiores a los costos, considerando los recursos que serán consumidos, como una manera de justificar la implementación del sistema, demostrando la factibilidad de los resultados esperados.

3.1.7 Obtener la aprobación del presupuesto

El director del proyecto y su equipo de trabajo elaboran un informe de solicitud de fondos para presentarlo formalmente a los responsables empresariales encargados de asegurar los recursos y fondos necesarios para el desarrollo del mismo. Aun después que el presupuesto ha sido aprobado, se deberán realizar reuniones periódicas, para comunicar los avances en el proyecto, y fomentar las expectativas y la seguridad en el presupuesto.

3.1.8 Desarrollar un plan de contingencias

Para la identificación y/o control de los riesgos potenciales se utiliza la evaluación de riesgos que identifica los riesgos potenciales y describe cómo evitarlos, además se elabora un plan de contingencia que registra el conjunto de tareas y pasos a realizar cuando los mismos se presentan, con la finalidad de minimizar sus consecuencias.

3.1.9 Suministrar capacitación

Su objetivo es adiestrar en las técnicas de trabajo en grupo y en la utilización de la metodología prevista para la construcción del DW a todos los miembros del proyecto, con el propósito de especificar la forma en que se acometerán las diferentes tareas y técnicas previstas para el cumplimiento de cada fase del proyecto y comprender cómo el DW se ajustará a la arquitectura de los sistemas empresariales.

3.2 Etapa 2 Modelaje

Esta etapa reviste vital importancia para el éxito del proyecto, ya que permite definir apropiadamente el modelo conceptual y el modelo físico de datos. El primer objetivo de esta etapa es determinar y documentar los requerimientos de todos los niveles de la organización.

Una buena táctica para comprender los principales aspectos a tener en cuenta en el diseño del DW es observar e investigar los procedimientos, los reportes que se emiten, los análisis de datos, etc., y determinar cómo influyen en el proceso de toma de decisiones.

De manera general, el proceso de modelaje de datos se fundamenta en tres niveles: el modelo conceptual, el modelo lógico y el modelo físico. Adicionalmente, para construir un DW, se precisa definir los principales componentes del depósito: componente de extracción, limpieza y agregación.

Objetivos:

- Refinar el alcance del sistema propuesto.
- Evaluar la tecnología actual y sus limitaciones.
- Identificar los componentes potenciales del sistema y sus requerimientos.
- Diseñar el esquema conceptual y físico del DW.
- Definir el proceso de carga, extracción, agregación y limpieza.
- Retroalimentar el proyecto a partir de los descubrimientos y mejoras obtenidas.

Pasos:

- 3.2.1. Estudio detallado de las fuentes de información.
- 3.2.2. Realizar el análisis empresarial.
- 3.2.3. Analizar la arquitectura tecnológica de la organización.
- 3.2.4. Desarrollo de un modelo conceptual.
 - 3.2.4.1. Identificar los tipos de entidades.
 - 3.2.4.2. Identificar las asociaciones.
 - 3.2.4.3. Identificar los atributos.
 - 3.2.4.4. Identificar el dominio.
 - 3.2.4.5. Estimar el tamaño de las fuentes de datos.
- 3.2.5. Diseño del DW.
 - 3.2.5.1. Identificar la tabla de hechos.
 - 3.2.5.2. Identificar las dimensiones requeridas.
 - 3.2.5.3. Determinar el esquema de agregación.
 - 3.2.5.4. Identificar las dimensiones cambiantes en el tiempo.
 - 3.2.5.5. Identificar las llaves.
- 3.2.6. Diseño físico del DW.
 - 3.2.6.1. Evaluar las técnicas de indexado.
 - 3.2.6.2. Evaluar el sistema de consultas.
 - 3.2.6.3. Definir y evaluar el sistema de acceso.
 - 3.2.6.4. Elaborar el esquema físico.
- 3.2.7. Definir los componentes del DW.
 - 3.2.7.1. Identificar los componentes de extracción.
 - 3.2.7.2. Identificar los componentes de limpieza.
 - 3.2.7.3. Identificar los componentes de agregación.
 - 3.2.7.4. Identificar el intérprete de consultas.
 - 3.2.7.5. Identificar los metadatos.
 - 3.2.7.6. Análisis de componentes.

3.2.1 Estudio detallado de las fuentes de información

El equipo de trabajo para realizar un efectivo análisis de la actividad empresarial debe:

- Desarrollar o revisar estándares (tanto técnicos como de procedimientos).
- Comunicar los estándares a todas las partes involucradas en el proyecto.
- Desarrollar un plan detallado para el proyecto.
- Definir los procedimientos para identificar y procesar los cambios que puedan producirse en la documentación estándar.

Con el objetivo de definir los estándares, es necesario realizar una revisión de la documentación, ya que la estructura de datos dentro del depósito probablemente sea muy diferente de los sistemas transaccionales.

Estos documentos serían los informes de la administración y aquellos suministrados por los propios departamentos de la organización como: informes anuales, planes empresariales, organigramas y manuales de operación, que deben mantenerse en un repositorio o biblioteca del proyecto. Otra fuente de información sería la proveniente de los materiales de investigación.

El equipo de trabajo debe esforzarse porque los nombres de tablas, dimensiones y atributos dentro del DW sean lo más amigables posibles, los cuales pudieran diferir bastante de las convenciones de nombres poseen los administradores de datos más convencionales.

3.2.2 Realizar el análisis empresarial

El análisis empresarial se utiliza para recolectar información sobre las principales actividades empresariales y los datos requeridos para su correcto funcionamiento.

Aspectos a tener en cuenta:

- Refinar el alcance empresarial.

En este punto lo mejor es establecer una clara diferenciación entre el alcance del proyecto y su desarrollo futuro, a fin de diseñar un mejor esquema de DW. Es aconsejable además, mantener el alcance dirigido a una primera y pequeña implementación que permita evaluar cualquier problema que se pueda presentar.

- Análisis de la cadena de valor empresarial¹⁴.

El análisis se centra en determinar las secuencias de actividades necesarias para satisfacer los principales objetivos empresariales. Desde el punto de vista del DW, este análisis ayudará a identificar qué puntos dentro del esquema del depósito deben estar relacionados, pues los mismos comparten dimensiones entre las tablas de hechos.

3.2.3 Analizar la arquitectura tecnológica de la organización

El objetivo fundamental de esta tarea es identificar las tecnologías de mayor rendimiento en el mercado actual y sus limitaciones, así como definir la arquitectura técnica que asumirá el proyecto, evaluando la aplicabilidad potencial de los paquetes de aplicaciones realizando un análisis de compra o reutilización.

Objetivo:

El propósito de esta tarea es obtener una comprensión de los sistemas en ejecución, así como de los elementos necesarios para la transición de estos sistemas al nuevo, referido a la evaluación y revisión del sistema actual incluyendo hardware, software e infraestructura de comunicación de todos aquellos sistemas que pueden interactuar con el sistema en desarrollo.

¹⁴ Conjunto de actividades requeridas para cumplimentar una actividad, sin considerar el mayor nivel de detalle.

3.2.4 Desarrollo del modelo conceptual

El modelo conceptual de datos debe definir las entidades y atributos necesarios para satisfacer las necesidades empresariales. Para elaborar este modelo es necesario realizar sesiones de trabajo en grupo y entrevistas de alta calidad, a fin de construir el modelo de entidad-relación¹⁵ y el diccionario de datos¹⁶, cuyos resultados sirven de preparación para el desarrollo del esquema físico de la base de datos, enfocando la atención en la estructura de la información: atributos, entidades y relaciones.

La identificación de todas las características de los datos que serán extraídos y cargados en el depósito (localización, calidad, cantidad, etc.) constituye un requisito preliminar para satisfacer el diseño del sistema e indicar los elementos a los que se enfrentará el proceso o componente de extracción, realizando una evaluación de la calidad de los datos, lo cual permitirá tomar importantes decisiones acerca de su limpieza y purificación.

3.2.4.1 Identificar los tipos de entidades

Una entidad es una representación de cualquier objeto o evento acerca del cual la empresa está interesada o que tiene determinado impacto en su desempeño.

En este punto hay que tener en cuenta que para el diseño y desarrollo de un DW el interés no se centra en la forma en que tradicionalmente se ejecuta un sistema informático transaccional, sino que el énfasis se centrará en la comprensión de qué tablas (representadas por tipos de entidades) se encuentran en los sistemas fuentes.

Recomendaciones:

- De ser posible obtenga los modelos de datos de los sistemas fuentes actuales que representen al sistema empresarial en funcionamiento, considerando qué puede suceder si algún cambio se produce y cómo éste impactará en el DW.
- Si los modelos de datos no están actualizados o no existen, utilice una herramienta CASE que permita realizar una reingeniería a las estructuras de datos.
- Cuando evalúe múltiples instancias de la misma fuente de datos, tome en cuenta los siguientes aspectos:
 - Ajustese a los objetivos empresariales.
 - Escoja de acuerdo con la proximidad que tenga a la fuente de datos original.
 - Completitud de la fuente (ejemplo, que todos los clientes estén listados en la tabla).
 - Calidad y precisión de los datos.
 - Analice las posibilidades de asociación de dicha fuente (por ejemplo, una tabla con los nombres de todos los clientes aparentemente puede ser buena, pero si no permite por su diseño tener una asociación con los datos demográficos, entonces no sería tan buena como otra que sí lo permita).

¹⁵ Es una representación gráfica de la estructura de los datos de relevancia para la empresa a fin de obtener una imagen de la estructura del sistema.

¹⁶ Contiene las definiciones de las entidades, asociaciones y atributos.

3.2.4.2 Identificar las asociaciones

Tal y como hemos planteado antes, el interés no está en un modelaje de datos como se realiza en los sistemas tradicionales, sino en comprender qué asociaciones entre las tablas (representadas por asociaciones entre entidades) están en las fuentes de datos.

Las asociaciones representan la interrelación entre entidades o entre entidades y subtipos de entidades, las cuales en última instancia representan el comportamiento empresarial y en muchos casos son utilizadas para captar algunas reglas empresariales. El entendimiento de estas asociaciones dará una idea de cuán complejo es el proceso de consulta al DW.

Recomendaciones:

- Investigue las asociaciones. Si el sistema fuente utiliza un SABDR, no presuponga que éste ya cumple con la integridad referencial.
- Preste especial atención a las relaciones de cardinalidad y la opcionalidad de las asociaciones, pues éstas pueden ser problemáticas para la consistencia del DW.

3.2.4.3 Identificar los atributos

Recuerde que estamos interesados en comprender qué campos (representados por atributos) dentro de las tablas se encuentran en los sistemas fuentes. Los atributos son necesarios para completar la definición de una entidad y deben describirla parcialmente.

3.2.4.4 Identificar el dominio

En esta tarea nos centraremos en qué tipo de valor o dominio poseen los atributos que se encuentran en los sistemas fuentes. Los dominios representan categorías que describen los posibles rangos de valores que un atributo puede tener.

3.2.4.5 Estimar el tamaño de las fuentes de datos

El objetivo fundamental es estimar el tamaño de las fuentes de datos, como insumo para determinar el tamaño de la base de datos que utilizará el DW. Por lo general el tamaño es expresado en términos de:

- Número de registros.
- Bytes por registros.
- Crecimiento estimado dado un período de tiempo.

Cada SABD tiene su propio conjunto de factores de estimación para identificar registros, espacio libre, bloques de control, etc., por lo que se debe revisar la documentación apropiada de acuerdo con el SABD que se vaya a utilizar para el DW.

Recomendaciones:

- El cálculo del tamaño sólo es válido para el proceso de extracción dirigido al DW. Por ejemplo, los sistemas fuentes pudieran tener ocho años de datos, pero en el depósito se puede planificar solo incluir los tres últimos años.
- En el proceso de extracción no se incluyen usualmente los atributos de texto de gran tamaño.
- En el cálculo del tamaño es beneficioso incluir las llaves (identificadores).

3.2.5 Diseño del DW

El diseño del sistema define la estructura del DW para que pueda ser utilizado por los usuarios empresariales. En términos prácticos el diseño de un DW según el modelaje propuesto por el esquema estrella¹⁷, equivale a determinar tabla de hechos, dimensiones, base de datos para el DW, herramientas OLAP, etc.

Objetivos:

- Refinar el alcance del sistema.
- Desarrollar el esquema de diseño del DW.
- Identificar los componentes potenciales del sistema.
- Identificar la tecnología existente y sus limitaciones.
- Determinar la tecnología que utilizará el proyecto.
- Establecer la documentación del sistema.

3.2.5.1 Identificar la tabla de hechos

Es la tabla central en una arquitectura estrella, la cual contiene todos los indicadores de ejecución (rendimiento) llaves (atributos, dimensión informacional, etc.); de manera general debe existir una tabla de hechos por cada proceso definido en el análisis de la cadena de valor empresarial.

En un diseño multidimensional posee una llave compuesta, donde cada componente de esta llave es una llave externa a una dimensión particular de la tabla. Sin embargo, determinar el tipo de esquema de diseño que mejor se adapte es un proceso iterativo que se mantendrá durante toda esta etapa, fundamentalmente en lo referente al refinamiento del diseño.

Recomendaciones:

- Mantenga siempre presente la cadena de valor empresarial.
- No trate de satisfacer muchos procesos empresariales con una sola tabla de hechos.
- Los hechos más valiosos son aquellos que son aditivos a través de todas las dimensiones.
- Nombre los atributos con denominaciones que sean fácilmente entendidas por los usuarios empresariales.
- No porque un atributo sea numérico, se convierte automáticamente en un hecho, por ejemplo, el número de factura de un producto es numérico, y de manera general, no se representa como una dimensión de producto.

Ante todo, hay que decidir qué se va a representar en la tabla de hechos, para lo cual es necesario escoger el gránulo. El gránulo de una tabla es la tupla individual que conforma la misma, en tal sentido es importante que el gránulo de la tabla de hechos esté correctamente definido, pues éste representa el menor nivel de detalle disponible en el depósito.

Sólo después de haber escogido correctamente el gránulo, se puede tener una discusión coherente de las dimensiones por considerar. En tal sentido, el gránulo de la tabla de hechos debe ser conocido antes de que se defina la dimensión. Por ejemplo, si el gránulo de la tabla es diario (como puede ser tomando como dimensión el tiempo) se sabe que se necesitará tener fechas (como 26-07-1998) en la dimensión tiempo y no la hora, pero si el gránulo de la dimensión es mensual, lo más probable es que no se necesite tanto nivel de detalle en la dimensión.

¹⁷ Ver Anexo 1.

3.2.5.2 *Identificar y conformar las dimensiones requeridas*

Esta tarea está encargada de identificar los elementos (dimensiones) que abarcará la tabla de hechos, siendo éstas las plataformas para examinar las restricciones disponibles.

En la medida que el esquema tome forma, busque dimensiones que puedan ser compartidas por diferentes tablas de hechos. Si alguna dimensión común es detectada, asegúrese de que son exactas y que pueden ser compartidas por las tablas de hechos que la poseen.

No normalice sus dimensiones, mejor agrúpelas en tablas planas de manera que tengan sentido para los usuarios empresariales y les permita examinar y revisar extensivamente los datos almacenados en las dimensiones.

Por ejemplo, la dimensión tiempo estructurada desde una perspectiva que permita optimizar la habilidad empresarial de seleccionar y analizar datos basados en el tiempo, podría tener los siguientes atributos:

- Fecha (en el formato admitido por la base de datos del SABD utilizado).
- Día de la semana (lunes a domingo).
- Día del mes (1-31).
- Semana del año (1-52).
- Mes del año (1-12).
- Nombre del mes (enero a diciembre).
- Año (1998).
- Año fiscal (1997).

3.2.5.3 *Determinar el esquema de agregación*

Los datos agregados representan resúmenes pre-calculados contruidos a partir de las tablas de hechos.

Dependiendo de los componentes del DW, es posible que se necesite definir un esquema para un agregado, en dependencia de las tablas de hechos y dimensiones identificadas. La creación de estos esquemas para los agregados es muy importante si se quiere que las consultas que se le realicen al depósito tengan un tiempo de respuesta significativamente rápido.

Es bastante común tener muchos esquemas de agregación, pues son muy necesarios para los analistas empresariales y el proceso de toma de decisiones, en tal sentido, es inevitable que en la medida que el depósito se desarrolla y utiliza se definan y creen nuevos esquemas de agregación.

Tome la debida cautela con la creación de agregados y no caiga en la tentación de querer controlarlo todo. Asegúrese que el agregado represente una cantidad considerablemente menor de filas que la tabla de hechos. Si los agregados no reducen el número de filas de la tabla de hechos por lo menos en un factor de 10, no construya el agregado.

3.2.5.4 *Identificar las dimensiones cambiantes en el tiempo*

Esta tarea identifica cómo la empresa desea que se manipulen los datos dentro de las dimensiones que cambian en el transcurso del tiempo, por tal razón es útil tener una medida de la calidad e integridad de las fuentes de datos antes de desarrollar esta tarea. Dado que en muchas ocasiones

el depósito se utiliza para evaluar datos históricos y realizar análisis de tendencias, es importante definir los cambios que desean tomarse en consideración.

Esta tarea está orientada a dar un tratamiento más detallado a las dimensiones, constituyendo un afinamiento en aquellos casos cuya duración es grande en el tiempo. Cuando se trabaja con períodos de tiempo grandes se presenta el problema de los cambios suaves en las dimensiones, caracterizados por dos fuentes fundamentales de cambios:

- Los cambios en los datos derivados de los sistemas fuentes.
- Los cambios en los valores que existen sólo en el DW (por ejemplo descripción del código de un sistema fuente).

Para ambos tipos de cambios se requiere analizar su efecto en la actualización y en la modificación, por ejemplo, en una actualización el cambio de nombre en un producto o en las regiones de ventas, puede provocar un cambio en las asociaciones dentro del sistema fuente, repercutiendo significativamente en el diseño de DW en dos aspectos:

- Es extremadamente difícil retroceder en el tiempo y acceder a las fuentes de los datos antiguos, presentándose problemas de lectura e interpretación.
- Es obligatorio usar las versiones antiguas de las dimensiones más importantes.

3.2.5.5 Identificar las llaves

Esta tarea identifica las llaves dentro del DW, lo cual se puede hacer en el paso de la definición de las tablas dimensionales o se puede realizar después en el paso de definición de los componentes. Las llaves necesitan ser identificadas para:

- Las tablas de hechos.
- Cada dimensión.
- Las dimensiones agregadas en el caso de que sean necesarias.

Si existe algún requerimiento de análisis a las fuentes de datos desde el depósito, entonces se debe reflexionar acerca del almacenamiento de identificadores de los sistemas fuentes.

Como regla general, una llave de una tabla de hechos es compuesta, es decir, está formada por múltiples campos de todas las dimensiones relacionadas. Si ésta no garantiza la unicidad es porque hay un mal diseño del modelo conceptual de las dimensiones y debe revisarse.

Recomendaciones:

- Utilice convenciones de nombres, por ejemplo utilice la palabra “key” para identificar los atributos llaves, lo cual permitirá a los usuarios finales y desarrolladores identificarlos fácilmente.
- Cuando defina una llave externa o foránea use el nombre del atributo donde la misma es llave.
- Tenga en cuenta que en muchos casos el equipo de desarrollo tendrá que administrar las llaves para los agregados, pues éstos no son obtenidos de los sistemas fuentes.

3.2.6 Diseño físico del DW

El diseño físico ha determinado el éxito o fracaso en muchos DW, el cual depende de un SABD particular o de una determinada configuración de hardware y suele ser necesario para alcanzar objetivos de rendimiento.

En este paso se abordan algunas pautas generales para el diseño del esquema físico del DW, a fin de decidir cómo ordenar la tabla de hechos en el disco, suponiendo que estamos en presencia de un modelo de datos completo y funcional que utiliza el esquema estrella.

El modelo físico determina la arquitectura del almacenamiento para el depósito, considerando la frecuencia de uso de determinados tipos de datos y las aplicaciones que apoyarán el DW. Actualmente, la mayoría de los paquetes de software diseñados para el modelaje de datos hacen la mayor parte del trabajo e incorporan elementos de rendimiento relacionados con el alto volumen de consultas y con el proceso de carga.

Los agregados usualmente suministran mejoras importantes en las características de rendimiento del DW, además, cuando se seleccionen las estrategias de acceso a los datos para el proceso de afinamiento del DW deben considerarse los efectos de ciertas técnicas de indexado, el optimizador de consultas y las mejoras en el proceso de navegación. Pero si aún así el rendimiento todavía está en tela de juicio, valore el multi-procesamiento para mejorar los tiempos de respuestas a las consultas.

3.2.6.1 Evaluar las técnicas de indexado

Se requiere evaluar varias técnicas de indexado para el DW, ya que las tablas dimensionales generalmente poseen muchos índices. En esta evaluación todos los elementos dependen en última instancia de los patrones de utilización de los datos. Los elementos a tener en cuenta son los siguientes:

- La tabla.
- Los atributos seleccionados para el índice.
- El tipo de índice utilizado.
- Si el índice es utilizado para particionar los datos.

Recomendaciones:

- Generalmente, las tablas de hechos poseen uno o dos grandes índices basados en una combinación de dimensiones llaves.
- Generalmente, una dimensión posee un índice construido para todos los atributos que pudieran examinarse. Esto significa que durante el proceso de carga el índice se destruye y antes de permitir el acceso a los datos se reconstruye.

3.2.6.2 Evaluar el sistema de consulta

Esta tarea identifica las bondades y limitaciones propias del optimizador de consultas del SABD que se esté utilizando, siendo el elemento básico el tiempo de respuesta del SABD a una consulta. Éste típicamente es una función del número físico de lecturas para cada fila, así como de la velocidad de procesamiento y de la cantidad de procesadores empleados.

En esta tarea la disyuntiva fundamental es analizar cómo el optimizador de consultas reduce el número de lecturas físicas necesarias para obtener el conjunto de filas requeridas por la consulta.

Los optimizadores más comunes crean índices de forma dinámica (lo cual implica que puedan construirlos en momentos inapropiados) o construyen índices cada vez que un usuario solicita una consulta, de acuerdo con el criterio de búsqueda. Si una búsqueda se realiza sobre una

significativa cantidad de filas entonces los optimizadores requieren de determinadas capacidades para cuando se desee incrementar su rendimiento, entre las que se encuentran:

- Escoger índices agrupados.
- Utilizar índices compartidos en vez de un índice a través de un *b-tree*.
- Seleccionar subconjuntos a través de filtros.
- Mantener para consultas adicionales, el conjunto de punteros encontrados que posean argumentos de búsquedas parecidos.

El elemento llave en última instancia es evaluar, para un particular atributo del depósito, el impacto de escoger uno u otro método de indexamiento.

3.2.6.3 Definir y evaluar el sistema de acceso

Esta tarea desarrolla el plan para la implementación de los índices a utilizar por el SABD del DW y describe la funcionalidad del sistema de consultas y las capacidades de navegación que serán utilizadas para el acceso a los datos.

Después de que las tablas físicas se han diseñado se deben incorporar reglas de control de acceso y seguridad al diseño físico que permitan ejecutar una administración confiable. De acuerdo con los requerimientos es recomendable no implementar niveles de seguridad sofisticados siempre que el entorno para acceder a la información sea sólo de lectura.

3.2.6.4 Elaborar el esquema físico

Esta tarea define el esquema físico de la base de datos para el depósito, generalmente desde un esquema estrella, incluyendo los agregados, los índices y las llaves. Idealmente toda la información para el desarrollo del esquema físico se debe captar a partir de una herramienta CASE que genere automáticamente el lenguaje de definición de datos (LDD).

3.2.7 Definir los componentes del DW

El objetivo de definir los requerimientos de información y las actividades propias de este paso en componentes es identificar los agrupamientos lógicos que faciliten la conformación del sistema de DW. Estos componentes se agrupan para lograr una mayor precisión en su construcción y para acelerar el desarrollo del sistema.

Recomendaciones:

- Es de gran ayuda a la hora de la creación de las interfaces entre los usuarios y los componentes del DW, una representación gráfica del listado de componentes (cartera de componentes¹⁸) adicional al documento textual.
- Una cartera de componentes bien definida constituye un excelente elemento de entrada para la planificación del proyecto, pues las tareas pueden ser más detalladas permitiendo ser más específicos en la asignación de recursos.

¹⁸ Una cartera de componentes suministra una información detallada de todos los datos, procesos, protocolos de comunicación y elementos tecnológicos de cada componente del sistema.

3.2.7.1 Identificar los componentes de extracción

Esta tarea identifica y define la cartera de componentes asociados con la extracción de datos. Dentro de los elementos que lo constituyen podemos citar:

- El proceso de consulta regular que se realiza de cada sistema fuente para alimentar el DW.
- La creación de las llaves.
- Las estrategias de consultas.
- La programación de la extracción.

3.2.7.2 Identificar los componentes de limpieza

Esta tarea identifica y define la cartera de componentes asociados con la transformación y purificación de los datos. Dentro de los elementos por considerar tenemos:

- La purificación.
- La transformación.

3.2.7.3 Identificar los componentes de agregación

Esta tarea identifica y define la cartera de componentes asociados con la definición de los agregados. Dentro de los elementos por tener en cuenta tenemos la carga al DW de:

- Los agregados.
- Las dimensiones.
- Las tablas de hechos.

3.2.7.4 Identificar el intérprete de consultas

Esta tarea identifica y define la cartera de componentes asociados con el intérprete de consultas. Dentro de los elementos a considerar podemos citar los siguientes:

- “*Drill across*” (procesamiento de la consulta de un esquema estrella a otro).
- “*Drill through*” (procesamiento de la consulta del depósito a los sistemas fuentes).
- Monitoreo de la consulta (obtener información acerca de la utilización de la consulta y de ser factible definir el agregado correspondiente y añadirlo al esquema del DW).
- Programación de las consultas (cómo procesar la consulta para obtener un tiempo de respuesta óptimo).

3.2.7.5 Identificar los metadatos

Esta tarea identifica y define la cartera de componentes asociados con los metadatos¹⁹. Dentro de los elementos por tener en cuenta tenemos:

- Las definiciones de datos.
- El esquema del DW (tablas de hechos y dimensiones).
- Los esquemas de diseño de los sistemas fuentes.
- Las vistas de las bases de datos.
- Las jerarquías dimensionales.

3.2.7.6 Análisis de componentes

En esta tarea se identifica la tecnología OLAP que pudiera ser utilizada para apoyar los procesos de toma de decisiones y la funcionalidad de los DSS, generalmente asociada con el análisis de los datos provenientes del depósito.

¹⁹ Información acerca de los datos que se encuentran en el DW.

Dentro de los elementos por considerar tenemos:

- Informes empresariales estándar.
- Aplicaciones de análisis de datos por tipo y herramienta.
- Análisis estadístico.
- Minería de datos.
- Administración de escenario (análisis de qué hacer si...).
- Generador del lenguaje de consultas.
- Ayuda en línea.

Debido a la gran cantidad de herramientas OLAP y de análisis de datos en el mercado, un aspecto importante a la hora de elegir es que sean fáciles de usar pues recuerde que están dirigidas a usuarios empresariales.

3.3 Etapa 3 Preparación e Implementación

Para probar el diseño lógico y físico realizado es recomendable elaborar un prototipo funcional, que deberá ser mostrado a todos los usuarios con el objetivo de detectar y fundamentar una estructura tecnológica que satisfaga el depósito (hardware, software, middleware, SABD, servidor, etc.); así como de las aplicaciones y programas necesarios que permitan poner a prueba la captura de datos, la integración, transferencia, carga y explotación.

El prototipo debe incorporar herramientas de usuario final que permitan interactuar, teniendo en cuenta qué verán y usarán, ya que de esa primera impresión depende que generen una buena imagen para valorar un producto que satisface sus requerimientos.

Objetivos:

- Poner el sistema en producción, con el mínimo de interrupción en las operaciones normales de la organización.
- Maximizar la efectividad del sistema, a fin de satisfacer las actividades empresariales.
- Identificar las potenciales mejoras al sistema, que pueden producirse como resultado de un proceso de retroalimentación.

Factores críticos de éxito:

- Preparar adecuadamente la plataforma de la implementación, el hardware, el software, las comunicaciones y las redes.
- Preparar y capacitar apropiadamente tanto a los usuarios finales como al equipo de apoyo operacional.

Recomendaciones:

- Cuando se está preparando la implementación, es muy importante considerar tanto los aspectos de capacitación como los de implementación física. En esta tarea el nuevo sistema es puesto en producción siguiendo un proceso de conversión de datos, capacitación del personal y estableciendo los procedimientos de producción necesarios para el buen funcionamiento del sistema.

- En cuanto a la forma de implementar puede ir desde una implementación rápida hasta una gradual o parcial. De manera general, es recomendable la implantación en paralelo para evitar al máximo las posibles interrupciones o mal funcionamiento de las actividades empresariales.
- Una buena táctica es involucrar con el éxito del proyecto a los usuarios finales o los designados para las pruebas representativas.
- Garantice que las opiniones y críticas sean canalizadas y que se produzca el adecuado proceso de retroalimentación.

En esta etapa se formalizan los pasos necesarios para que el entorno del nuevo sistema diseñado comience a funcionar.

3.3.1. Preparación

- 3.3.1.1 Establecer el equipo de apoyo al sistema
- 3.3.1.2 Planificar el proyecto de implementación
- 3.3.1.3 Establecer el plan de capacitación
- 3.3.1.4 Establecer el entorno de implementación
- 3.3.1.5 Ejecutar la carga preliminar del DW

3.3.2. Implementación

- 3.3.2.1 Ajustar hardware y software
- 3.3.2.2 Suspender el procesamiento actual
- 4.3.2.3 Ejecutar rutinas de conversión

3.3.3. Evaluación

- 3.3.3.1 Dirigir el proceso de implementación
- 3.3.3.2 Retroalimentación con los resultados obtenidos

3.3.1 Preparación

En este paso se trata de ejecutar el plan de implementación del sistema, teniendo como objetivos:

- Planificar y organizar la instalación, prueba y operación inicial.
- Preparar el entorno de operación del sistema.

3.3.1.1 Establecer el equipo de apoyo al sistema

En esta tarea se designa el equipo responsable de la administración, mantenimiento y mejoramiento gradual del sistema, tomando en consideración los objetivos generales propuestos por el desarrollo del DW, los cuales deben satisfacerse con la implementación. Estas personas deben poseer las habilidades requeridas para satisfacer las necesidades de apoyo al sistema propuesto, obtenidas vía capacitación y experiencia práctica.

3.3.1.2 Planificar el proyecto de implementación

La planificación de la implementación tiene como propósito definir las tareas, recursos, la capacitación y la programación detallada para implementar el sistema.

Este plan debe incluir:

- El programa de cómo se establecerá la implementación (gradual, prototipo funcional, pruebas pilotos, pruebas en paralelo, etc.).
- Los costos asociados con la implementación.
- La actualización del plan de contingencias.

- El plan de conversión de datos.
- La documentación de la retroalimentación.

3.3.1.3 Establecer el plan de capacitación

Una implementación exitosa requiere experiencia en las operaciones que un DW demanda de la TI empleada. En tal sentido, los materiales y la documentación técnica del sistema deben entregarse a la comunidad usuaria para solventar sus dudas en el empleo del sistema propuesto, basado en las necesidades expresadas en el plan de implementación.

Con este fin se deben considerar el número de personas por capacitar y la frecuencia de la capacitación, las condiciones y facilidades para desarrollarlo, incluyendo no sólo el aula de clases, sino el hardware, el software y el entorno simulado en el cual se llevará a cabo.

3.3.1.4 Establecer el entorno de implementación

El entorno o ambiente del sistema debe ser identificado antes de la instalación del mismo, de hecho los requerimientos del entorno son evaluados y planificados en la etapa de diseño, tomando como base la documentación del sistema operacional.

3.3.1.5 Ejecutar la carga preliminar del DW

A partir de los componentes del DW que fueron construidos u obtenidos vía “*outsourcing*” se poblará el DW. Este proceso de carga es una prueba inicial para comprobar el funcionamiento de los componentes y del SABD del DW, este proceso culminará en la etapa de operación. Esta prueba sirve para que los administradores de sistemas monitoreen de manera cercana todo el proceso para comprobar si se satisfacen los requerimientos.

3.3.2 Implementación

Este paso busca la implementación completa del sistema de DW, teniendo entre sus principales objetivos:

- Establecer el sistema en producción.
- Retroalimentar a la administración del proyecto con las sugerencias, las críticas y los descubrimientos.
- Asegurar el software y hardware, así como las facilidades para la prueba del sistema.

3.3.2.1 Ajustar hardware y software

Es probable que se requiera cambiar o actualizar el sistema operativo, el SABD o el software de comunicaciones para implementar el DW.

3.3.2.2 Suspender el procesamiento actual

Durante el proceso de implementación algunas o todas las actividades empresariales pueden ser interrumpidas, por tal motivo deberán ser cuidadosamente planificadas para garantizar el funcionamiento de la organización, debiendo minimizarse dicho período de suspensión de operaciones.

La conversión de los datos del sistema actual al propuesto es una tarea delicada porque las operaciones actuales deben suspenderse antes de instalar el nuevo sistema. Las aplicaciones

actuales pueden quedar fuera de operación hasta que los reemplazos necesarios para el funcionamiento del nuevo sistema sean operacionales. De aquí la importancia de que el proceso de conversión de datos sea ejecutado lo más rápidamente posible.

3.3.2.3 Ejecutar rutinas de conversión

Los datos deben ser adaptados o convertidos para que puedan ser utilizados por el nuevo sistema, ya que utilizan rutinas y procedimientos que no necesariamente son compatibles con los sistemas actuales. Es sumamente importante para evitar riesgos y minimizar sus consecuencias establecer de antemano las rutinas de recuperación por si ocurre un fallo en el proceso de conversión.

El resultado del esfuerzo en el proceso de conversión de datos debe ser cuidadosamente revisado antes de que sea puesto en operación. Una buena estrategia es dirigir métodos para el chequeo de la precisión y la integridad de las conversiones realizadas.

3.3.3 Evaluación

El sistema implementado o en implementación gradual es probado para asegurar que satisface la funcionalidad exigida, los requerimientos de rendimiento y las expectativas de los usuarios.

Objetivos:

- Garantizar que el sistema reúna los requerimientos especificados así como las expectativas esperadas.
- Retroalimentar con las críticas, sugerencias y descubrimientos.

3.3.3.1 *Dirigir el proceso de implementación*

Significa probar el sistema utilizando como base de comparación la funcionalidad definida para el mismo. Las pruebas de aceptación incluyen un conjunto de ejercicios y operaciones sobre el sistema que suministran el grado de satisfacción usuaria y la necesaria retroalimentación.

3.3.3.2 *Retroalimentación con los resultados obtenidos*

Como resultado de la implementación pueden ocurrir descubrimientos que tengan cierto impacto en el sistema, los que deberán considerarse en el proceso de retroalimentación.

3.4 Etapa 4 Operación

Esta etapa se centra en la operación y mejoramiento del sistema implementado, capturando los requerimientos que impliquen aumento de su calidad, según las siguientes actividades:

- Seleccionar un conjunto de métricas para la operación del sistema propuesto.
- Recoger todos los elementos que permitan un proceso de retroalimentación efectivo.
- Analizar la calidad del funcionamiento del sistema.

Para ello es necesario obtener toda la información que sea posible sobre las experiencias durante la implementación del sistema, incluyendo información acerca de actividades tales como consumo de recursos, duración de las tareas, lecciones aprendidas, etc., a fin de intensificar el proceso de mejoramiento del proyecto.

Por lo general, en los proyectos informáticos esta etapa está descuidada, de hecho se ve como la etapa de asignación de culpas sobre aquellos detalles que no se ejecutaron correctamente, cuando debiera considerarse como un proceso de aprendizaje a partir de la experiencia adquirida durante las etapas anteriores que permita mejorar el mantenimiento y afinamiento del DW.

Objetivos:

- Evaluar y recomendar mejoras a la administración del proyecto.
- Proveer ayuda adicional durante un período de tiempo hasta que el sistema se estabilice.
- Identificar las mejoras potenciales al sistema, para que sean consideradas por la administración del proyecto en futuras implementaciones o afinamiento del sistema.
- Examinar el rendimiento del proyecto y su calidad final, tomando como parámetros los objetivos planificados.
- Afinamiento del DW.

Factores críticos de éxito:

- Disponibilidad de información, fundamentalmente referente al alcance y al consumo de tiempo y recursos durante el proyecto.
- Retroalimentar los descubrimientos, críticas y sugerencias hechas al proyecto durante su desarrollo.
- Apoyo de la máxima gerencia, en lo referido a fondos y recursos requeridos para dirigir el programa de evaluación y mejoramiento de la calidad.
- Apoyo de los expertos de la comunidad empresarial, para que le sean suministrados los datos y elementos necesarios para la retroalimentación del proyecto.

Pasos:

- 3.4.1 Suministrar asistencia a los usuarios para el funcionamiento del sistema
- 3.4.2 Suministrar asistencia operacional
- 3.4.3 Análisis del rendimiento del sistema
- 3.4.4 Afinamiento del DW
- 3.4.5 Evaluación post-implementación

3.4.1 Suministrar asistencia a los usuarios para el funcionamiento del sistema

Como es conocido, con la instalación de todo nuevo sistema se generan muchas dudas acerca del proceso de manipulación y operación del mismo; por ello, los usuarios del nuevo sistema requieren de un considerable apoyo durante las primeras semanas de operación.

Este apoyo está a cargo del administrador de las aplicaciones de análisis de datos y del responsable de capacitar al equipo de desarrollo del proyecto, los cuales además de adiestrar a los usuarios tienen la responsabilidad de captar toda la información necesaria que permita retroalimentar el funcionamiento del DW, así como identificar y resolver los problemas asociados con el software, ya que en ocasiones el software implementado puede tener errores.

Aspectos a tener en cuenta:

- Suministrar el hardware y software necesario.

Incluye las facilidades requeridas para las operaciones del sistema en cuando a comunicaciones (protocolos y topología de red), así como a la arquitectura de desarrollo empleada (ejemplo: cliente/servidor y toda la infraestructura que ello conlleva).

- Suministrar el mantenimiento y la capacitación necesaria.

Incluye la ayuda necesaria en la operación del nuevo sistema y en la programación de adiestramiento personalizado al personal involucrado con la operación del proyecto.

3.4.2 Suministrar asistencia operacional

Un DW introduce inevitablemente cambios en el entorno operacional que generan confusión y error, por ello el administrador del depósito debe estar muy alerta con la organización del mantenimiento, para informar cualquier cambio que impacte notablemente a los usuarios a fin de brindarles oportunamente el apoyo necesario.

3.4.3 Análisis del rendimiento del sistema

En la medida que el sistema se utiliza, el rendimiento se debe medir y monitorear con el propósito de afinar el sistema. La mayor parte del monitoreo para el DW, se realiza al intérprete de consultas, el cual es analizado por el administrador del depósito.

Aspectos a tener en cuenta:

- Métricas para analizar el rendimiento del sistema.
- Evaluación del rendimiento del intérprete de consultas.

3.4.4 Afinamiento del DW

Este paso involucra el análisis del rendimiento del DW y el análisis del sistema actual, lo cual puede generar la modificación del diseño del esquema estrella con el objetivo de mejorar los tiempos de respuesta de las consultas y el proceso de generación de informes para los usuarios finales.

3.4.5 Evaluación post-implementación

La puesta en funcionamiento no significa el final de la implementación, ya que es necesario revisar la distribución de datos y las entregas de información. Para evaluar la implementación y los resultados alcanzados resulta beneficioso responder:

En el corto plazo

¿Qué podría ser mejor?

¿Cómo el proyecto satisfizo a la gerencia y cómo la gerencia ayudó al proyecto?

¿De qué manera trabajaron las partes involucradas?

¿Cuáles son los beneficios inmediatos para los usuarios?

¿Se seleccionaron las herramientas correctas?

En el mediano plazo

¿Fueron considerados los aspectos sustantivos en las áreas?

¿Fueron escogidos los departamentos determinantes?

¿Qué resultados se alcanzaron?

¿Cómo fueron publicados estos resultados?

A largo plazo

¿El DW aportó beneficios y valor agregado real a la empresa?

¿Significó el DW una ventaja competitiva para la empresa?

¿Cómo se puede valorar el retorno de la inversión?

¿Qué conseguimos que no esperábamos?

Resumen de los pasos para la implementación de un depósito de datos.

Etapa 1 Fundamentación

- | | |
|--|---|
| 1. Definir las fronteras del sistema. | 6. Analizar costo beneficio. |
| 2. Identificar los objetivos empresariales del DW. | 7. Obtener la aprobación del presupuesto. |
| 3. Identificar las consultas potenciales. | 8. Desarrollar un plan de contingencias. |
| 4. Identificar las fuentes de datos potenciales. | 9. Suministrar capacitación. |
| 5. Pronosticar los volúmenes iniciales. | |

Etapa 2 Modelaje

- | | |
|--|---|
| 1. Estudio detallado de las fuentes de información. | 6. Diseño físico del DW.
Evaluar las técnicas de indexado.
Evaluar el sistema de consultas.
Definir y evaluar el sistema de acceso.
Elaborar el esquema físico. |
| 2. Realizar el análisis empresarial. | |
| 3. Analizar la arquitectura tecnológica de la organización. | |
| 4. Desarrollo de un modelo conceptual.
Identificar los tipos de entidades.
Identificar las asociaciones.
Identificar los atributos.
Identificar el dominio.
Estimar el tamaño de las fuentes de datos. | 7. Definir los componentes del DW.
Identificar los componentes de extracción.
Identificar los componentes de limpieza.
Identificar los componentes de agregación.
Identificar el intérprete de consultas.
Identificar los metadatos.
Análisis de componentes. |
| 5. Diseño del DW.
Identificar la tabla de hechos.
Identificar las dimensiones requeridas.
Determinar el esquema de agregación.
Identificar las dimensiones cambiantes en el tiempo.
Identificar las llaves. | |

Etapa 3 Preparación e Implementación

- | | |
|--|---|
| 1. Preparación
Establecer el equipo de apoyo al sistema
Planificar el proyecto de implementación
Establecer el plan de capacitación
Establecer el entorno de implementación
Ejecutar la carga preliminar del DW | 2. Implementación
Ajustar hardware y software
Suspender el procesamiento actual
Ejecutar rutinas de conversión |
| | 3. Evaluación
Dirigir el proceso de implementación
Retroalimentación con los resultados obtenidos |

Etapa 4 Operación

- | | |
|---|---|
| 1. Suministrar asistencia a los usuarios para el funcionamiento del sistema | 3. Análisis del rendimiento del sistema |
| 2. Suministrar asistencia operacional | 4. Afinamiento del DW |
| | 5. Evaluación post-implementación |

Anexo No. 1 Esquema estrella

El objetivo de este esquema es el diseño de la estructura de datos para el DW. Es también utilizado para predecir el rendimiento y características del depósito así como la naturaleza de las consultas que el mismo puede satisfacer. Este diagrama puede ser almacenado en una herramienta CASE, el cual incluye el esquema de la tabla de hechos, las dimensiones, sus asociaciones, identificadores y atributos.

Este esquema plantea un enfoque de modelado que posee un sólo objeto conectado con varios objetos de manera radial. El objeto en el centro de la estrella se denomina *tabla de hechos* y los objetos conectados se denominan *tablas de dimensión*. Un esquema de estrella sencillo consta de una tabla de hechos y varias tablas de dimensión. Los esquemas estrellas complejos poseen cientos de tablas de hechos y de dimensión. La tabla de hechos contiene las mediciones básicas de las empresas con millones de filas. Las tablas de dimensión contienen atributos de las empresas que se emplean como criterios de búsqueda y son relativamente pequeñas.

En la Figura No.1 se muestra el diseño de un esquema de DW según el modelaje propuesto por el esquema estrella, que a modo de ejemplo presenta un modelo de ventas que consta de cuatro dimensiones, cuya estructura es tomada del artículo [McGuff 1996].

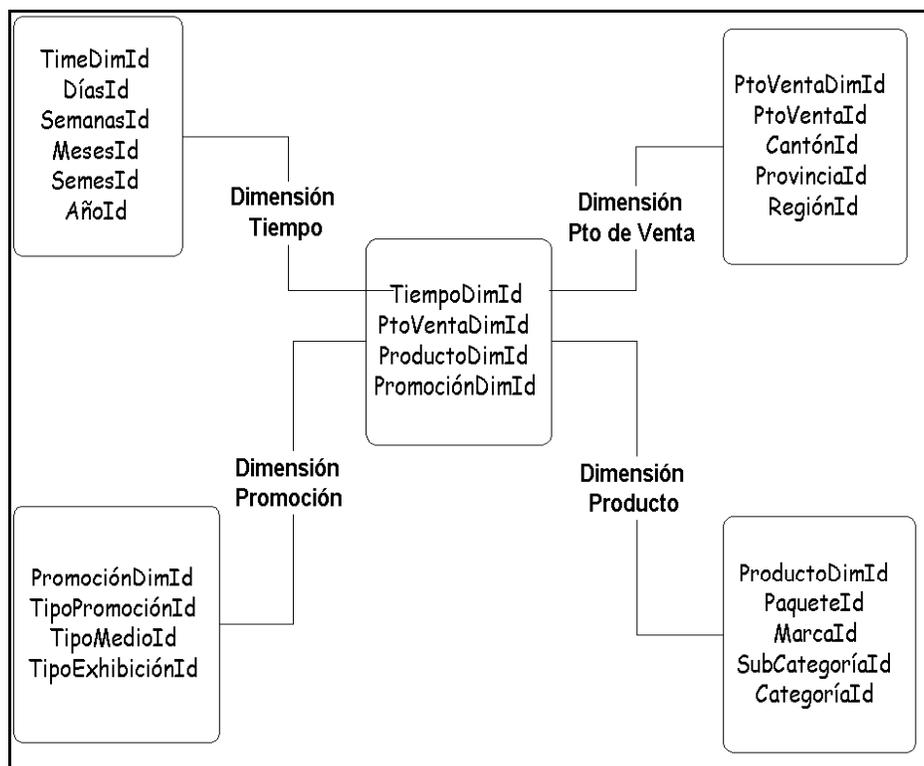


Figura No. 1 Esquema estrella

Anexo No. 2 Glosario

Análisis multidimensional de datos

Análisis simultáneo de múltiples dimensiones de datos.

Administración de datos distribuidos

Es la arquitectura que permite que programas de aplicación o usuarios de un sistema tengan acceso a datos almacenados en sistemas remotos.

Base de datos multidimensional

Base de datos diseñada alrededor de un conjunto de dimensiones que usa análisis multidimensional.

Base de datos operacional

Base de datos que apoya sistemas de software que están soportando normalmente las operaciones de una organización. También denominadas bases de datos OLTP.

Cadena de valor empresarial

Es la secuencia de procesos que describe el movimiento o trayectoria de los productos o servicios que la organización empresarial trabaja desde su inicio hasta su venta final.

CASE

Siglas de Computer Aided Software Engineering (ingeniería de software apoyada con computadoras). Una metodología de desarrollo de software soportada por herramientas basadas en computación para el apoyo de tareas de análisis, diseño y desarrollo.

Cliente Gordo

Está en una arquitectura cliente/servidor, donde la estación de trabajo del cliente administra el procesamiento de la información (lógica empresarial) así como la interfaz gráfica del usuario.

Data warehouse (depósito de datos)

Estructura organizativa de los datos dentro de un almacén específicamente dirigida a satisfacer el procesamiento de consultas y de análisis de los datos.

Datos de herencia

Fuentes de datos históricos para el depósito de datos, los cuales con frecuencia en medios fuera de línea.

Datos derivados

Datos producidos por cálculos o procesos que aplica el depósito de datos a los datos provenientes de las fuentes. Los datos derivados sólo residen en el depósito de datos y éste los usa para precalcular varios valores que se requieren para responder a solicitudes frecuentes.

Dimensión

Los datos dentro del depósito de datos se organizan de manera tal de almacenar grandes volúmenes de datos históricos y operacionales utilizando múltiples dimensiones para establecer categorías. Una de las más importantes es el tiempo.

Dimensiones cambiantes en el tiempo

Es la tendencia de los registros dimensionales, específicamente las dimensiones producto y clientes, de cambiar gradualmente en el tiempo.

Drill across

Es la actividad de solicitar datos de una o más tablas de hechos en un mismo reporte.

Drill down

Es el acto de profundizar en el conjunto respuesta a una determinada consulta, lo cual significa elevar el nivel de detalle. Generalmente implica añadir nuevas filas al reporte.

Drill Up

Es el acto de consolidar o resumir el conjunto respuesta a una determinada consulta, lo cual significa disminuir el nivel detalle. Generalmente significa eliminar filas al reporte.

Entornos distribuidos

Constituyen los estándares entornos computacionales desarrollados por la Open Software Foundation (OSF) que proveen interoperabilidad y portabilidad a través de sistemas distribuidos.

Esquema estrella

Es un paradigma de modelado que tiene en medio un objeto único conectado a varios objetos en forma radial. Este esquema refleja la visión del usuario en una consulta empresarial: hechos tales como ventas, compensaciones, pagos o facturas son calificadas por una o más dimensiones (por mes, producto, región geográfica, etc.). El objeto en el centro del esquema estrella se llama tabla de hechos y los que se conectan desde las periferias se llaman tabla dimensión.

Grano o gránulo

Es el significado de un registro en la tabla de hechos. Su determinación constituye uno de los pasos para diseñar el esquema estrella.

Granularidad

Término que se usa en los depósitos de datos para expresar el nivel de detalle. A más alto nivel de granularidad, más bajo nivel, más bajo nivel de detalle (más alto nivel de abstracción).

Integración de datos

La actividad de combinar datos de diversas fuentes en el depósito de datos a fin de presentar una colección única de datos para el usuario del depósito.

Integridad referencial

Condición de obligatoriedad dentro de los depósitos de datos, donde todas las claves en la tabla de hechos son claves foráneas que apuntan a cada una de las tablas dimensión.

Lenguaje de definición de datos (DDL)

Parte del lenguaje de consulta estructurado (SQL) que consiste de los mandatos responsables de la creación o eliminación de los objetos bases de datos.

Lenguaje de manipulación de datos (DML)

Parte del lenguaje de consulta estructurado (SQL) que consiste de los mandatos responsables de la manipulación de los datos, por ejemplo SELECT, DELETE, UPDATE, INSERT.

Limpieza y refinamiento de datos

La actividad de eliminar errores en las fuentes de datos antes de intentar cargarlas en el depósito de datos o el mercado de datos.

Mercado de datos (data mart)

Una implementación del depósito de datos, pero con un dominio más pequeño y restringido, que sirve a un departamento o a una parte de la organización. Por lo general, una organización posee varios mercados de datos.

Middleware

El término común que se aplica al software que intercambia información en forma transparente entre aplicaciones y bases de datos. El middleware es un mecanismo de conexión abstracta entre el software de aplicación y la base de datos, y oculta al programador de aplicaciones los elementos específicos que dependen de la implementación.

Minería de datos

El proceso eficiente de descubrir información valiosa para la organización dentro de grandes colecciones de datos.

Multiprocesadores simétricos (MPS)

Divide el trabajo de múltiples procesadores en una sola CPU.

OLTP

Siglas de ON-Line Transaction Processing (procesamiento analítico en línea). Término que se usa para definir cualquier sistema de software que reúne los datos usando las transacciones (en el momento en que ocurren) entre la fuente de datos y la base de datos.

Orientado a temas

Clasificado por el interés empresarial.

Procesamiento analítico en línea (OLAP)

Tecnología de análisis multidimensional de datos y capacidad de generación de reportes.

Procesamiento analítico

Modalidad de verificación de análisis de datos, usualmente utilizado en el análisis multidimensional de los datos.

Segmentar

Es la actividad de dividir un dato en dos o más elementos pequeños. Se usa para separar una tabla única de la fuente de datos en dos o más tablas dentro del depósito de datos. También se aplica para dividir las filas de una tabla en dos o más tablas con la misma estructura, pero de manera que cada una contenga un subconjunto de las filas.

Sistemas de apoyo a la toma de decisiones (DSS)

Sistemas automatizados de aplicaciones que ayudan a la organización a tomar decisiones relacionadas con la empresa.

Sistema de información ejecutiva (EIS)

Término que se utiliza para aquellos sistemas de consulta y generación de reportes que ejecutan adiciones y resúmenes directamente en los datos operacionales, algunas veces guardando datos resumidos y proporcionan determinada ayuda a los encargados de la toma de decisiones.

Sistemas legados

Sistema operacional generalmente dedicado a las funciones empresariales de la organización. Puede no ser un sistema transaccional o relacional. Usualmente reside en un mainframe.

Snapshot (instantáneas)

Un tipo de tabla de hecho que representa el estado de una determinada característica, entidad o dimensión al final de un período de tiempo.

Bibliografía

- [Alur 1995] Alur, N: "Missing Links in the Data Warehousing."
Database Programming & Design. Vol. 8, No. 9, Págs. 21-23, Septiembre, 1995.
- [Alur 1996] Alur, N: "Data Warehousing & Parallel Technology."
Database Programming & Design. Vol. 9, No. 1, Págs. 23-26, Enero, 1996.
- [Avalos 1997] Avalos, Rafael y Jiménez, Beatriz: "Depósito de datos".
Seminario de Investigación I, Maestría en Computación, ITCR. Agosto de 1997.
- [Barquin 1997] Barquin, Ramon et al: "Planning and Designing the Data Warehouse"
Prentice Hall. 1997.
- [Berson 1997] Berson, Alex et al: "Data Warehousing, Data Mining & OLAP."
McGraw Hill. 1997.
- [Bischoff 1994] Bischoff, Joyce: "Achieving Warehouse Success."
Database Programming & Design. Vol. 7, No. 7, Págs. 27-33, Julio, 1994.
- [Bischoff 1996] Bischoff, J., Yevich, R: "Building More Than a Data Warehouse."
Database Programming & Design. Vol. 9, No. 9, Págs. 27-31, Septiembre, 1996.
- [Carlson 1997] Carlson, Bob: "Secrets of success."
Data Warehouse Information Center. [<http://pwp.startnetinc.com/larryg/index.html>]. Marzo, 1997
- [Celko 1995] Celko, Joe, McDonald, Jackie: "Don't Warehouse Dirty Data"
[<http://www.datamation.com>] 1995
- [Demarest 1996] Demarest, Marc: "Evaluating Data Warehousing Technologies: Oracle 7.1."
Data Warehouse Information Center. [<http://pwp.startnetinc.com/larryg/index.html>]. 1996.
- [Demarest 1997] Demarest, Marc: "The Politics of Data Warehousing."
Data Warehouse Information Center. [<http://pwp.startnetinc.com/larryg/index.html>]. Junio, 1997.
- [Devlin 1997] Devlin, Barry: "Data Warehouse from Architecture to Implementation".
Addison-Wesley. 1997.
- [Dwic 1996] Data Warehouse Information Center: "Datamation's One minute Guide to Data Warehousing." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1996.
- [Dwop 1996] Data Warehouse Operation: "Data Warehousing Operation."
Data Warehousing Operation. [<http://www.dwo.bull.com>]. 1996.
- [Eckerson 1995] Eckerson, Wayne W: "Building the Legacy Systems of Tomorrow."
Data Warehouse Information Center. [<http://pwp.startnetinc.com/larryg/index.html>]. 1995.

- [Edelstein 1995]** Edelstein, Herb: "Technology Analysis: Faster Data Warehouse." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1995.
- [Finkelstein 1996]** Finkelstein, R: "Sybase IQ: Expressly for the Warehouse." *Database Programming & Design*. Vol. 9, No. 12, Págs. 47-49, Diciembre, 1996.
- [Freed 1996]** Freed, Anita J: "Building Success Into Your Data Warehouse Project." *DCI's Data Warehouse World*. [<http://www.dciexcpo.com/datawhse/>]. 1996.
- [Giordano 1993]** Giordano, Robert: "The Information "Where?" house." *Database Programming & Design*. Vol. 6, No. 9, Págs. 54-58, Septiembre, 1993.
- [Greenfield 1996]** Greenfield, Larry: "Data Warehousing." [<http://www.datamation.com>]. 1996.
- [Hackathorn 1995]** Hackathorn, Richard: "Data Warehouse Energizes your Enterprise." [<http://www.datamation.com>]. 1995.
- [Harjinder 1996]** Harjinder, S. Gill y Prakash C. Rao: "Data Warehousing. La integración de información para la mejor toma de decisiones." Prentice Hall. 1996.
- [Hernández 1997]** Hernández, Edgar y Martínez, Luis: "Cliente/servidor". Editado y publicado por Rho-Sigma, S.A. a nombre del Club de Investigación Tecnológica. Diciembre 1997.
- [Horrocks 1995]** Horrocks, Chris: "Data Warehouse: An executive Perspective" *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. Noviembre, 1995.
- [Horrocks 1996]** Horrocks, Chris: "Making the warehouse work." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. Junio, 1996.
- [Hufford 1996]** Hufford, Duane: "Data Warehouse Quality Part I." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1996.
- [Inmon 1991]** Inmon, W. H: "The Age of the Warehouse." *Database Programming & Design*. Vol. 4, No. 8, Págs. 68-69, Agosto, 1991.
- [Inmon 1992]** Inmon, W. H: "Dawn Of The Warehouse." *Database Programming & Design*. Vol. 5, No. 12, Págs. 76-77, Diciembre, 1992.
- [Inmon 1992a]** Inmon, W. H: "EIS and the Warehouse." *Database Programming & Design*. Vol. 5, No. 11, Págs. 70-71, Noviembre, 1992.
- [Inmon 1995]** Inmon, W.H: "What is a Data Warehouse?." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1995.
- [Inmon 1996a]** Inmon, W. H: "What is Data Mart?." *D2K, Inc*. [<http://www.d2k.com/d2k/library2.htm>]. 1996.
- [Kimball 1996]** Kimball, Ralph: "Letting The Users Sleep, Part 1." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. Diciembre, 1996.

- [Kimball 1997]** Kimball, Ralph: "Letting the users sleep, Part 2." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. Enero, 1997.
- [Kulkarni 1997]** Kulkarni, J. y King, R: "Business Intelligent System and Data Mining." *SAS Institute*. [<http://www.sas.com/sitemap/index.html>]. 1997.
- [Lambert 1995]** Lambert, Bob: "What do you need to Know to Succeed." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1995.
- [Lambert 1996]** Lambert, Bob: "Break Old Habits To define Data Warehousing." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1996.
- [Madsen 1996]** Madsen, M: "Warehouse Design in The Aggregate." *Database Programming & Design*. Vol. 9, No. 7, Págs. 45-52, Julio, 1996.
- [Mansell 1996]** Mansell-Lewis, Emma: "Data Warehouse: Size isn't everythings." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1996.
- [McElreath 1996]** McElreath, Jack: "Data Warehouse: An Architecture Perspective." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1996.
- [McGuff 1996]** McGuff, Frank: "Data Modelling for Data Warehouse." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. Octubre, 1996.
- [Meredith 1996]** Meredith, M.E., Khader, A: "Divide and Aggregate: Design Large Warehouse." *Database Programming & Design*. Vol. 9, No. 6, Págs. 25-30, 25-30, Junio, 1996.
- [Moriarty 1995]** Moriarty, Terry: "Avoiding The Trauma." *Database Programming & Design*. Vol. 8, No. 11, Págs. 53-55, Noviembre, 1995.
- [Moriarty 1995a]** Moriarty, Terry: "A Data Warehouse Primer." *Database Programming & Design*. Vol. 8, No. 7, Págs. 57-59, Julio, 1995.
- [Moriarty 1995b]** Moriarty, Terry: "Modeling Data Warehouse." *Database Programming & Design*. Vol. 8, No. 8, Págs. 61-65, Agosto, 1995.
- [Moriarty 1995c]** Moriarty, Terry: "Part of Whole." *Database Programming & Design*. Vol. 8, No. 10, Págs. 53-55, Octubre, 1995.
- [Moriarty 1996]** Moriarty, Terry: "What Every Good System Does?." *Database Programming & Design*. Vol. 9, No. 3, Págs. 57-59, Marzo, 1996.
- [Moriarty 1996a]** Moriarty, T., Greenwood, R: "Data's Quest From Source Query." *Database Programming & Design*. Vol. 9, No. 10, Págs. 78-81, Octubre, 1996.
- [Moriarty 1996b]** Moriarty, T. Moore, A: "Data Warehouse Must Read." *Database Programming & Design*. Vol. 9, No. 12, Págs. 61-63, Diciembre, 1996.

[**Moriarty 1996c**] Moriarty, T., Mandracchia, C: "Heart of Warehouse." *Database Programming & Design*. Vol. 9, No. 11, Págs. 70-71, Noviembre, 1996.

[**Moriarty 1996d**] Moriarty, Terry: "Architecting for Evolution." *Database Programming & Design*. Vol. 9, No. 6, Págs. 58-63, Junio, 1996.

[**Mundy 1995**] Mundy, Joy: "From The Ground Up: Building a Data Warehouse." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1995.

[**Parsaye 1995a**] Parsaye, K. y Chignell, M: "Quality Unbound." *Database Programming & Design*. Vol. 8, No. 1, Págs. 38, 40-42, Enero, 1995.

[**Parsaye 1996a**] Parsaye, Kamram: "Surveying Decision Support." *Database Programming & Design*. Vol. 9, No. 8, Págs. 27,28-30, 32-33, Abril, 1996.

[**Parsaye 1997**] Parsaye, Kamran: "Bridging the Gap." *Database Programming & Design*. Vol. 10, No. 2, Págs. 31-36, Febrero, 1997.

[**Pearson 1996**] Pearson, D: "More Than Meets The Eye." *Database Programming & Design*. Vol. 9, No. 9, Págs. 58-63, Septiembre, 1996.

[**Poe 1994**] Poe, Vidett: "Guidelines for Warehouse Development." *Database Programming & Design*. Vol. 7, No. 9, Págs. 60-64, Septiembre, 1994.

[**Poe 1995**] Poe, Vidette: "Data warehouse is NOT infrastructure." *Database Programming & Design*. Vol. 8, No. 7, Págs. 24-31, Julio, 1995.

[**Quinlan 1996**] Quinlan, T: "Report From The Trenches." *Database Programming & Design*. Vol. 9, No. 12, Págs. 37-38, Diciembre, 1996.

[**Rpi 1997**] Rpi Inc: "Frequently Asked Questions." *RPI*. [<http://www.rpi.edu/~vanepa2/faq.htm>]. 1997.

[**Sas 1997c**] SAS Institute Inc: "The SAS Data Warehouse." *SAS Institute Inc*.1997.

[**Sas 1997d**] SAS Institute Inc: "Data Warehousing. A Blueprint for Success." *SAS Institute Inc*. 1997.

[**Sas 1997e**] SAS Institute Inc: "Data Warehousing. A Checklist for Success." *SAS Institute Inc*. 1997.

[**Sas 1997f**] SAS Institute Inc: "Data Warehousing: Understanding Its Role in Bussiness Management Architecture." *SAS Institute Inc*.1997.

[**Schur 1994**] Schur, Stephen G: "Reengineering: Destination WorkFlow." *Database Programming & Design*. Vol. 7, No. 12, Págs. 52-58, Diciembre, 1994.

[**Simon 1995**] Simon, Alan: "I Want a Data Warehouse So, What is It Again?." *Database Programming & Design*. Vol. 8, No. 12, Págs. 26-31, Diciembre, 1995.

- [**Simon 1996**] Simon, A: "Beyond The Warehouse." *Database Programming & Design*. Vol. 9, No. 13, Págs. 43-46, Diciembre, 1996.
- [**Stephen 1997**] Stephen, Amy: "Turning Data Into Decisions." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1997.
- [**Stodder 1996**] Stodder, D: "Business of Data: An Interview With M. Saylor." *Database Programming & Design*. Vol. 9, No. 10, Págs. s2-s8, Octubre, 1996.
- [**Stodder 1997**] Stodder, David: "The Secret is Out." *Database Programming & Design*. Vol. 10, No. 5, Págs. 7-8, Mayo, 1997.
- [**Strehlo 1997**] Strehlo, Kevin: "Data Warehousing: Avoid Planned Obsolescence." <Http://www.datamation.com>. 1997.
- [**Thomson 1996**] Thomson, Doug: "In Search of the Ideal." *Database Programming & Design*. Vol. 9, No. 6, Págs. 62-66, Junio, 1996.
- [**Thorntwaite 1997**] Thorntwaite, Warren, et al: "Industry BrainStorm: What's Hot in Data Warehousing?." *Data Warehouse Information Center*. [<http://pwp.startnetinc.com/larryg/index.html>]. 1997.
- [**White 1993**] White, Colin: "Why a Data Warehouse." *Database Programming & Design*. Vol. 6, No. 9, Págs. 27-29, Septiembre, 1993.
- [**White 1994**] White, Colin: "Let The Replication Battle Begin." *Database Programming & Design*. Vol. 7, No. 5, Págs. 21-24, Mayo, 1994.
- [**White 1995**] White, Colin: "The Key to Data Warehouse." *Database Programming & Design*. Vol. 8, No. 2, Págs. 23-25, Febrero, 1995.
- [**White 1996**] White, Colin: "Data Warehouse: What's in a Name." *Database Programming & Design*. Vol. 9, No. 3, Págs. 53-54, Marzo, 1996.
- [**Youngworth 1994**] Youngworth, Paul: "Build a Data Warehouse Solution." *Databased Advisor*. Vol. 12, No. 7, Págs. 48-49, Julio, 1994.
- [**Youngworth 1994a**] Youngworth, Paul: "Data Warehouse Meets Schreiber Foods's Needs." *Databased Advisor*. Vol. 12, No. 7, Págs. 50-51, Julio, 1994.
- [**Zibitsker 1996**] Zibitsker, B., Saigalov, G: "Effective MPP Performance Management." *Database Programming & Design*. Vol. 9, No. 10, Págs. s11-s16, Octubre, 1996.
- [**Zornes 1996**] Zornes, Aaron: "A Taxonomy of Corporate Data Warehouse." *DCI's Data Warehouse World*. [<http://www.dciexpo.com/datawhse/>]. 1996.
- [**Zyga 1997**] Zyga Corporation: "CEM for Data Warehouse". [<http://www.zyga.com>]. 1997.

Capítulo 1 Fundamentos de un depósito de datos	1
1.1 Usos y beneficios de un depósito de datos	6
1.2 Factores Críticos de Éxito	8
1.3 Razones para construir un depósito de datos	9
1.4 Pautas por considerar en el desarrollo de un depósito de datos	11
1.5 Mercados de datos	14
Capítulo 2 Arquitecturas de depósitos de datos	16
2.1 Arquitectura básica	17
2.2 Arquitectura genérica	19
2.3 Arquitectura por niveles	23
2.1.3.1 Arquitectura de dos niveles	23
2.1.3.2 Arquitectura de tres niveles	25
2.4 Arquitectura por bloques	26
2.5 Arquitectura basada en el modelo empresarial	31
Capítulo 3 Proceso de implementación de un depósito de datos.....	36
3.1 Etapa 1 Fundamentación	37
3.1.1 <i>Definir las fronteras del sistema</i>	37
3.1.2 <i>Identificar los objetivos empresariales del DW</i>	38
3.1.3 <i>Identificar las consultas potenciales</i>	38
3.1.4 <i>Identificar las fuentes de datos potenciales</i>	38
3.1.5 <i>Pronosticar los volúmenes iniciales</i>	38
3.1.6 <i>Analizar costo beneficio</i>	39
3.1.7 <i>Obtener la aprobación del presupuesto</i>	39
3.1.8 <i>Desarrollar un plan de contingencias</i>	39
3.1.9 <i>Suministrar capacitación</i>	39
3.2 Etapa 2 Modelaje	39
3.2.1 <i>Estudio detallado de las fuentes de información</i>	40
3.2.2 <i>Realizar el análisis empresarial</i>	41
3.2.3 <i>Analizar la arquitectura tecnológica de la organización</i>	41
3.2.4 <i>Desarrollo del modelo conceptual</i>	42
3.2.4.1 Identificar los tipos de entidades	42
3.2.4.2 Identificar las asociaciones	43
3.2.4.3 Identificar los atributos	43
3.2.4.4 Identificar el dominio	43
3.2.4.5 Estimar el tamaño de las fuentes de datos	43
3.2.5 <i>Diseño del DW</i>	44
3.2.5.1 Identificar la tabla de hechos	44
3.2.5.2 Identificar y conformar las dimensiones requeridas	45
3.2.5.3 Determinar el esquema de agregación	45
3.2.5.4 Identificar las dimensiones cambiantes en el tiempo	45
3.2.5.5 Identificar las llaves	46
3.2.6 <i>Diseño físico del DW</i>	46
3.2.6.1 Evaluar las técnicas de indexado	47
3.2.6.2 Evaluar el sistema de consulta	47
3.2.6.3 Definir y evaluar el sistema de acceso	48
3.2.6.4 Elaborar el esquema físico	48
3.2.7 <i>Definir los componentes del DW</i>	48
3.2.7.1 Identificar los componentes de extracción	49
3.2.7.2 Identificar los componentes de limpieza	49
3.2.7.3 Identificar los componentes de agregación	49
3.2.7.4 Identificar el intérprete de consultas	49
3.2.7.5 Identificar los metadatos	49
3.2.7.6 Análisis de componentes	49
3.3 Etapa 3 Preparación e Implementación	50

3.3.1 Preparación.....	51
3.3.1.1 Establecer el equipo de apoyo al sistema	51
3.3.1.2 Planificar el proyecto de implementación	51
3.3.1.3 Establecer el plan de capacitación	52
3.3.1.4 Establecer el entorno de implementación.....	52
3.3.1.5 Ejecutar la carga preliminar del DW	52
3.3.2 Implementación	52
3.3.2.1 Ajustar hardware y software	52
3.3.2.2 Suspender el procesamiento actual	52
3.3.2.3 Ejecutar rutinas de conversión	53
3.3.3 Evaluación.....	53
3.3.3.1 Dirigir el proceso de implementación	53
3.3.3.2 Retroalimentación con los resultados obtenidos	53
3.4 Etapa 4 Operación.....	53
3.4.1 Suministrar asistencia a los usuarios para el funcionamiento del sistema	54
3.4.2 Suministrar asistencia operacional	55
3.4.3 Análisis del rendimiento del sistema.....	55
3.4.4 Afinamiento del DW	55
3.4.5 Evaluación post-implementación.....	55
Resumen de los pasos para la implementación de un depósito de datos.	57
Anexo No. 1 Esquema estrella	58
Anexo No. 2 Glosario.....	59
Bibliografía.....	62