

Primeros pasos hacia la Inteligencia de Negocios:

Almacenes de datos (DataWareHouse), Herramientas ETL(PDI), Procesamiento analítico en línea.

Introducción

Una solución de Business Intelligence parte de los sistemas de origen de información de una organización o entidad, donde se encuentran diferentes Bases de Datos, ficheros de texto, etc, sobre los que suele ser necesario aplicar una transformación estructural para optimizar su proceso analítico, importante para la toma de decisiones de los directivos de la entidad u organización. Para ello se realiza una fase de extracción, transformación y carga (ETL) de datos. Esta etapa suele apoyarse en un almacén intermedio, llamado ODS, que actúa como pasarela entre los sistemas fuente y los sistemas destino (generalmente un Datawarehouse), y cuyo principal objetivo consiste en evitar la saturación de los servidores funcionales de la organización, o sea, evitar la carga en los servidores, producto de las diferentes encuestas a los diferentes orígenes de información en busca de datos e información útil para la toma de decisiones. La información resultante, ya unificada, depurada y consolidada, se almacena en un Datawarehouse corporativo, que puede servir como base para la construcción de distintos Datamarts departamentales. Estos Datamarts se caracterizan por poseer la estructura óptima para el análisis de los datos de esa área de la empresa, ya sea mediante bases de datos transaccionales (OLTP) o mediante bases de datos analíticas (OLAP). Los datos albergados en el Datawarehouse o en cada Datamart se explotan utilizando herramientas de análisis, reporting, alertas, etcétera.

Business Intelligences el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP / OLAP, alertas...) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.[1]

Los principales componentes de orígenes de datos en el Business Intelligence que existen en la actualidad son Datamart y DataWarehouse.

Los componentes del BI se diferencian de los sistemas operacionales en que están optimizados para preguntar y divulgar sobre datos. Esto significa que, en un DataWarehouse, los datos están desnormalizados para apoyar consultas de alto rendimiento, mientras que en los sistemas operacionales suelen encontrarse normalizados para apoyar operaciones continuas de inserción, modificación y eliminación de datos. En este sentido, los procesos ETL, que nutren los sistemas BI, tienen que traducir de uno o varios sistemas operacionales normalizados e independientes a un único sistema desnormalizado, cuyos datos estén completamente integrados.

En definitiva, una solución BI completa permite: Observar qué está ocurriendo, comprender por qué ocurre, predecir qué ocurriría, colaborar qué debería hacer el equipo, decidir qué camino se debe seguir.[1]

DATAWAREHOUSE Y DATAMART

Datawarehouse

Un Datawarehouse es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. [2]

En el funcionamiento de un almacén de los datos es muy importante la integración y separación de los datos. La integración de los datos, no es más que los datos provenientes de bases de datos distribuidos por las diferentes unidades de la organización y que con frecuencia tendrán diferentes estructuras. Se debe facilitar una descripción global y un análisis comprensivo de toda la organización en el almacén de datos. La separación de los datos consiste en separar los datos usados en operaciones diarias, de los datos del Datawarehouse para los propósitos de divulgación, análisis y ayuda en la toma de decisiones y para operaciones de control. Ambos tipos de datos no deben coincidir en la misma base de datos, ya que obedecen a objetivos muy distintos y podrían entorpecerse entre sí.

Un almacén de datos es una colección de datos que está formada por variables (hechos o facts) y dimensiones (dimensions). Las dimensiones son los elementos para ubicar datos que participan en el análisis y las variables los valores que se desean analizar.

Tabla de hechos

En un Datawarehouse, una Tabla de Hechos (o tabla fact) es la tabla central de un esquema dimensional y contiene los valores de las medidas de negocio. Cada medida se toma mediante la intersección de las dimensiones que la definen, dichas dimensiones estarán reflejadas en sus correspondientes tablas de dimensiones que rodearán la tabla de hechos y estarán relacionadas con ella.

Las medidas del negocio (hechos)

Las medidas más útiles para incluir en una Tabla de hechos son los aditivos, es decir, aquellas medidas que pueden ser sumadas como por ejemplo, la cantidad de producto vendido, los costos de producción o el dinero obtenido por las ventas; son medidas numéricas que pueden calcularse con la suma de varias cantidades de la tabla. En consecuencia, por lo general los hechos a almacenar en una tabla de hechos van a ser casi siempre valores numéricos, enteros o reales.

Cardinalidad de la tabla de hechos

Las tablas de Hechos pueden contener un gran número de filas, a veces cientos de millones de registros cuando contienen uno o más años de la historia de una gran organización, esta cardinalidad estará acotada superiormente por la cardinalidad de las tablas dimensionales, Por ejemplo, si se tiene una tabla de hechos "TH" de tres dimensiones D1, D2 y D3, el número máximo de elementos que tendrá la tabla de hechos TH será:

$\text{Card (TH)} = \text{Card (D1)} \times \text{Card (D2)} \times \text{Card (D3)}$, Donde 'Card(x)' es la cardinalidad de la tabla 'x'

Granularidad

Una característica importante que define a una tabla de hechos es el nivel de granularidad de los datos que en ella se almacenan, entendiéndose por 'granularidad' el nivel de detalle de dichos datos, es decir, la granularidad de la tabla de hechos representa el nivel más atómico por el cual se definen los datos. Por ejemplo, no es lo mismo contar el tiempo por horas (grano fino) que por semanas (grano grueso); o en el caso de los productos, se puede considerar cada variante de un mismo artículo como un producto (por ejemplo, en una empresa textil, cada talla y color de pantalón podría ser un producto) o agrupar todos los artículos de una misma familia considerándolos como un único producto (por ejemplo, el producto pantalón genérico). Como se puede observar, la granularidad afecta a la cardinalidad, tanto de las dimensiones como de la Tabla de hechos, a mayor granularidad (grano más fino) mayor será el número de registros final de la Tabla de Hechos.

Cuando la granularidad es mayor, es frecuente que se desee disponer de subtotales parciales, es decir, si tenemos una Tabla de hechos con las ventas por días, podría interesar disponer de los totales semanales o mensuales, estos datos se pueden calcular haciendo sumas parciales, pero es frecuente añadir a la Tabla de hechos registros donde se almacenan dichos cálculos para no tener que repetirlos cada vez que se requieran y mejorar así el rendimiento de la aplicación. En este caso se dispondrá en la misma Tabla de hechos de datos de grano fino y de grano más grueso aumentando aún más la cardinalidad de la tabla.

Agregación

La agregación es un proceso de cálculo por el cual se resumen los datos de los registros de detalle, esta operación consiste normalmente en el cálculo de totales dando lugar a medidas de grano grueso. Cuando se resumen los datos, el detalle ya no está directamente disponible para el analista, ya que este se elimina de la tabla de hechos. Esta operación se realiza típicamente con los datos más antiguos de la empresa con la finalidad de seguir disponiendo de dicha información (aunque sea resumida) para poder eliminar registros obsoletos de la tabla de hechos para liberar espacio.

Esquema de estrella es la arquitectura de Datawarehouse más simple. En este diseño del almacén de datos, la tabla de variables (hechos) está rodeada por dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP. El motivo por dejar de mantener las tablas en el modelo relacional y permitir el almacenamiento de información redundante, es optimizar el tiempo de respuesta de base datos y dar información a un usuario en menos tiempo posible.

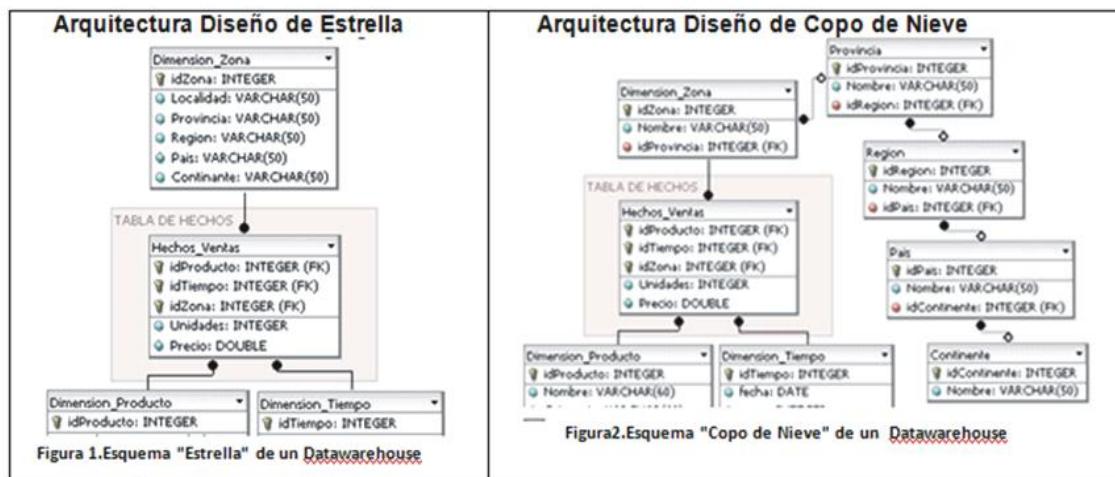
La clave primaria de una tabla de hechos está formada por todas las columnas que corresponden a las dimensiones (como por ejemplo ID_CLIENTE, ID_ARTICULO, ID_TIEMPO). Las columnas que contienen los datos numéricos no forman parte de la clave primaria, porque están agregadas en los informes (ejemplos de este tipo de columnas: CANTIDAD, PRECIO, VENTA_TOTAL, IVA). [5]

Esquema en copo de nieve o bola de nieve, es una variedad más compleja del esquema de estrella. El afinamiento está orientado a facilitar mantenimiento de dimensiones. Existen

varios usos del esquema en bola de nieve, el más común es cuando las tablas de dimensiones están muy grandes o complejas y es muy difícil representar los datos en esquema de estrella.

Por ejemplo, si una tabla dimensional de los clientes (CUSTOMERS) contiene un millón de filas, sería una idea buena crear una tabla con grupos de clientes (CUSTOMER_GROUPS) y mover los datos comunes para cada grupo de clientes a esta tabla. El tamaño de estas dos tablas será mucho menor que de una tabla no normalizada con todos los datos de clientes.

El problema es que para extraer datos de las tablas en esquema de copo de nieve, a veces hay que vincular muchas tablas en las sentencias SQL que puede llegar a ser muy complejo y difícil para mantener. [5]



Arquitectura Diseño de Constelación de Hech

Para cada esquema estrella o esquema del Copo de Nieve en el almacén de datos es posible construir un esquema de constelación de hechos. Este esquema es más complejo que las otras arquitecturas debido al hecho de que contiene múltiples tablas de hechos. Con esta solución las tablas de dimensiones pueden estar compartidas entre más de una tabla de los hechos. El esquema de constelación de hechos tiene mucha flexibilidad y este hecho es su grande virtud. Sin embargo, el problema es que cuando el número de tablas vinculadas aumenta, la arquitectura puede llegar a ser muy compleja y difícil para mantener.

Los almacenes de datos contienen a menudo datos y parámetros introducidos manualmente para mantener configuración solicitada. Los almacenes de datos contienen información que se subdivide a veces en unidades lógicas más pequeñas, llamadas centros comerciales (Data marts).

Almacén de Datos, al contrario de OLTP, es un sistema que debe dar una respuesta a cada pregunta con respeto al funcionamiento de la empresa. Es una base de datos que almacena información, la cual se construye a partir de datos básicos extraídos de otras bases de datos operacionales. El objetivo de estos Datawarehouse es ayudar a la toma de decisiones con la capacidad de realizar en tiempo real, análisis multidimensionales. Esta información tiene que ser accesible rápidamente, pero no es tan importante que sea lo más actual posible y al nivel de detalle muy bajo. Normalmente un almacén de datos se carga diariamente durante la

noche cuando los sistemas fuente no están sobrecargados. El objeto mayor de la arquitectura de un almacén de datos es guardar datos históricos y agregados.[5]

Datamart

Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer de la estructura óptima de datos para analizar la información al detalle, desde todas las perspectivas que afecten a los procesos de dicho departamento. Un Datamart puede ser alimentado desde los datos de un Datawarehouse, o integrar por sí mismo un compendio de distintas fuentes de información.

Por tanto, para crear el Datamart de un área funcional de la empresa, es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio Datawarehouse, o sobre una base de datos OLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento. [3]

Entre las características de un Datamart se destacan:

- Usuarios limitados.
- Área específica.
- Tiene un propósito específico.
- Tiene una función de apoyo.

Aportes de los Datawarehouse

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

Ventajas

- Los almacenes de datos hacen más fácil el acceso a una gran variedad de datos a los usuarios finales.

- Facilitan el funcionamiento de las aplicaciones de los sistemas de apoyo a la decisión tales como informes de tendencia, por ejemplo: obtener los items con la mayoría de las ventas en un área en particular dentro de los últimos dos años; informes de excepción, informes que muestran los resultados reales frente a los objetivos planteados a priori.

- Los almacenes de datos pueden trabajar en conjunto y, por lo tanto, aumentar el valor operacional de las aplicaciones empresariales, en especial la gestión de relaciones con clientes.

Desventajas

Utilizar almacenes de datos también tiene algunos inconvenientes, algunos de ellos son:

- A lo largo de su vida los almacenes de datos pueden suponer altos costos. El almacén de datos no suele ser estático. Los costos de mantenimiento son elevados.

- Los almacenes de datos se pueden quedar obsoletos relativamente pronto.

- A veces, ante una petición de información estos devuelven una información subóptima, que también supone una pérdida para la organización.

- A menudo existe una delgada línea entre los almacenes de datos y los sistemas operacionales. Hay que determinar qué funcionalidades de estos se pueden aprovechar y cuáles se deben implementar en el Datawarehouse, resultaría costoso implementar operaciones no necesarias o dejar de implementar alguna que sí vaya a necesitarse.

Por último, destacar que para comprender íntegramente el concepto de Datawarehouse, es importante entender cuál es el proceso de construcción del mismo, denominado ETL, a partir de los sistemas operacionales de una entidad u organización.

HERRAMIENTA ETL

ETL son las siglas en inglés de Extraer, Transformar y Cargar (Extract, Transform and Load). Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, transformarlos y cargarlos en otra base de datos, datamart, o datawarehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos y con la extracción se convierten los datos a un formato preparado para iniciar el proceso de transformación.

La fase de transformación aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Algunas fuentes de datos requerirán alguna pequeña manipulación de los datos.

La fase de carga es el momento en el cual los datos de la fase anterior (transformación) son cargados en el sistema de destino. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. Los Almacenes de Datos mantienen un historial de los registros de

manera que se pueda hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

Algunas de las herramientas ETL existentes son:

Ab Initio Beneti BITool – ETL Software CloverETL CognosDecisionstream (IBM) Data Integrator (herramienta de Sap Business Objects) ETI*Extract (ahora llamada EtiSolution) IBM Websphere DataStage (antes AscentialDataStage) Microsoft Integration Services Oracle Warehouse Builder	Kettle (ahora llamado Pentaho Data Integration) Scriptella Open Source ETL Tool. Talend Open Studio. Jitterbit
---	--

Las herramientas ETL no siempre se utilizan en entornos de construcción de un almacén de datos, sino que pueden ser útiles para multitud de propósitos, como por ejemplo: migración de datos entre diferentes aplicaciones, sincronización entre diferentes sistemas, consolidación de datos (Sistemas con grandes volúmenes de datos que son consolidados en sistemas paralelos para mantener históricos o para procesos de borrado en los sistemas originales), entre otros.

A continuación se realiza una breve explicación de unas de las herramientas ETL: PDI (Pentaho Data Integration anteriormente llamado Kettle).

El Pentaho Data Integration (PDI) está formado por un conjunto de herramientas, cada una con un propósito específico.

- Spoon: es la herramienta gráfica que nos permite el diseño de las transformaciones y trabajos del sistema ETL de Pentaho Data Integration, también conocido como Kettle. Incluye opciones para previsualizar y testear los elementos desarrollados. Es la principal herramienta de trabajo de PDI y con la que construiremos y validaremos nuestros procesos ETL.

- Pan: es la herramienta que nos permite la ejecución de las transformaciones diseñadas en spoon (bien desde un fichero o desde el repositorio). Nos permite desde la línea de comandos preparar la ejecución mediante scripts.

- Kitchen: similar a Pan, pero para ejecutar los trabajos o jobs.

- Carte: es un pequeño servidor web que permite la ejecución remota de transformaciones y jobs.

Transformaciones

La transformación es el elemento básico de diseño de los procesos ETL en PDI. Una transformación se compone de pasos o steps, que están enlazados entre sí a través de los saltos o hops. Los pasos son el elemento más pequeño dentro de las transformaciones y constituyen el elemento a través del cual fluye la información entre los diferentes pasos (siempre es la salida de un paso y la entrada de otro). En la Figura 3 se muestra que en el primer paso se están recuperando registros de una tabla de la base de datos, y los registros

recuperados van siendo transmitidos a los siguientes pasos a través del salto, y se van realizando operaciones sobre los datos con los diferentes pasos incluidos.

Tenemos un amplio repertorio disponible de pasos que nos permiten abordar casi cualquier necesidad en el diseño de los procesos de integración de datos. Los pasos están agrupados por categorías y cada uno de ellos está diseñado para cumplir una función determinada. Cada paso tiene una ventana de configuración específica, donde se determinan los elementos a tratar y su forma de comportamiento.

Una transformación no es ningún programa ni un ejecutable, simplemente es un conjunto de metadatos en XML que le indican al motor de PDI las acciones a realizar.

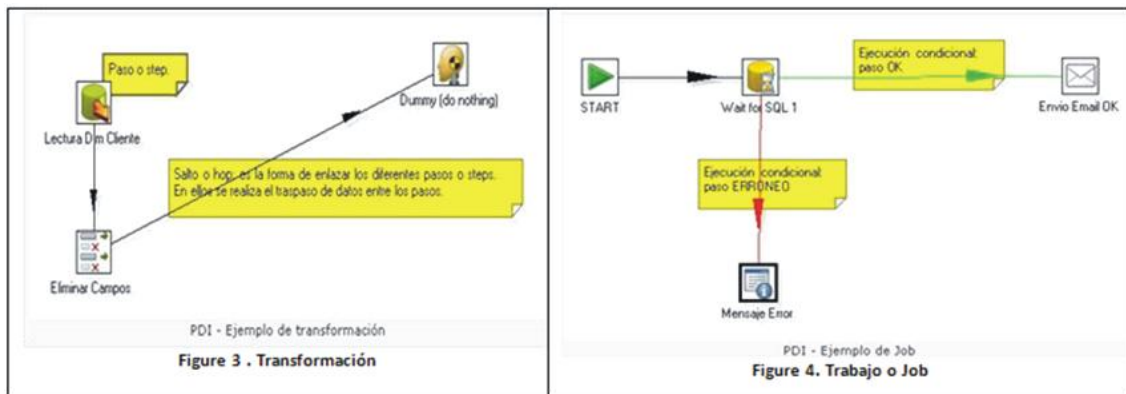
Trabajo o Job

Un Trabajo o Job, es similar al concepto de proceso. Un proceso es un conjunto sencillo o complejo de tareas con el objetivo de realizar una acción determinada. En los trabajos podemos utilizar pasos específicos (que son diferentes a los disponibles en las transformaciones) como para recibir un fichero vía ftp, mandar un email, ejecutar un comando, etc. Además, podemos ejecutar una o varias transformaciones de las que hayamos diseñado y orquestar una secuencia de ejecución de ellas. Los trabajos estarían en un nivel superior a las transformaciones.

Los saltos o hops entre los componentes de un job, indican el orden de ejecución de cada uno de ellos (no empezando la ejecución del elemento siguiente hasta que el anterior no ha concluido). El paso de un componente del job a otro también puede ser condicional, según si el resultado de ejecución ha sido correcto o no.

Al igual que las transformaciones, un job no es ningún programa, es también un conjunto de metadatos en XML, que le describen al motor de PDI la forma de realizar las diferentes acciones.

Periódicamente, se importan datos al almacén de datos de los distintos sistemas de planeamiento de recursos de la entidad (ERP) y de otros sistemas de software relacionados con el negocio para la transformación posterior. Es práctica común normalizar los datos antes de combinarlos en el almacén de datos mediante herramientas ETL. Estas herramientas leen los datos primarios (a menudo bases de datos OLTP de un negocio), realizan el proceso de transformación al almacén de datos (filtración, adaptación, cambios de formato, etc.) y finalmente escriben los datos transformados en él.



PROCESAMIENTO ANALÍTICO EN LINEA (CUBOS OLAP)

Los datos almacenados en los DataWarehouse son inútiles si no se explotan de forma correcta, por lo que la capa de presentación tiene vital importancia.

OLAP es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo Inteligencia Empresarial (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales (OLTP). Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.

La razón de usar OLAP para las consultas es la velocidad de respuesta. Una base de datos relacional almacena entidades en tablas discretas si han sido normalizadas. Esta estructura es buena en un sistema OLTP, pero para las complejas consultas multitabla es relativamente lenta. Un modelo mejor para búsquedas (aunque peor desde el punto de vista operativo) es una base de datos multidimensional. La principal característica que potencia a OLAP, es que es lo más rápido a la hora de ejecutar sentencias SQL de tipo SELECT, en contraposición con OLTP que es la mejor opción para operaciones de tipo INSERT, UPDATE Y DELETE.

Funcionalidad

En la base de cualquier sistema OLAP se encuentra el concepto de CUBO OLAP (también llamado cubo multidimensional o hipercubo). Se compone de hechos numéricos llamados medidas que se clasifican por dimensiones.

Tipos de sistemas OLAP

Tradicionalmente, los sistemas OLAP se clasifican según las siguientes categorías:

ROLAP

Implementación OLAP que almacena los datos en un motor relacional. Típicamente, los datos son detallados, evitando las agregaciones y las tablas se encuentran normalizadas. Los esquemas más comunes sobre los que se trabaja son estrella ó copo de nieve, aunque es posible trabajar sobre cualquier base de datos relacional. La arquitectura está compuesta por un servidor de banco de datos relacional y el motor OLAP se encuentra en un servidor dedicado. La principal ventaja de esta arquitectura es que permite el análisis de una enorme

cantidad de datos. Típicamente ROLAP se usa, para largos conjuntos de datos que no son frecuentemente buscados, tales como datos históricos de los años más recientes.

MOLAP

Esta implementación OLAP almacena los datos en una base de datos multidimensional. Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado. Estos valores precalculados o agregaciones son la base de las ganancias de desempeño de este sistema. Algunos sistemas utilizan técnicas de compresión de datos para disminuir el espacio de almacenamiento en disco debido a los valores precalculados. En general este método, es muy apropiado para el uso de búsquedas que requieran rápida respuesta.

HOLAP (Hybrid OLAP)

Almacena algunos datos en un motor relacional y otros en una base de datos multidimensional.

HOLAP combina atributos de MOLAP y ROLAP, la agregación de datos es almacenada en una estructura multidimensional usada por MOLAP, y la base de datos fuentes, en una base de datos relacional. Para procedimiento de búsqueda que accesan datos sumariados, HOLAP es equivalente a MOLAP, por el contrario si estos procesos accasaran datos fuentes, estos deben de buscar los datos en la base de datos relacional y esto no es tan rápido comparado a si los datos estuvieran almacenados en una estructura MOLAP.

Comparación

Cada sistema OLAP tiene ciertos beneficios (aunque existe desacuerdo acerca de las características específicas de los beneficios entre los proveedores). Algunas implementaciones MOLAP son propensas a la "explosión" de la base de datos; este fenómeno provoca la necesidad de grandes cantidades de espacio de almacenamiento para el uso de una base de datos MOLAP cuando se dan ciertas condiciones: elevado número de dimensiones, resultados precalculados y escasos datos multidimensionales. Las técnicas habituales de atenuación de la explosión de la base de datos no son todo lo eficientes que sería deseable.

Por lo general, MOLAP ofrece mejor rendimiento debido a la especializada indexación y a las optimizaciones de almacenamiento. MOLAP también necesita menos espacio de almacenamiento en comparación con los especializados ROLAP porque su almacenamiento especializado normalmente incluye técnicas de compresión. ROLAP es generalmente más escalable. Sin embargo, el gran volumen de preprocesamiento es difícil de implementar eficientemente por lo que con frecuencia se omite; por tanto, el rendimiento de una consulta ROLAP puede verse afectado. Desde la aparición de ROLAP van apareciendo nuevas versiones de bases de datos preparadas para realizar cálculos, las funciones especializadas que se pueden utilizar tienen más limitaciones. HOLAP (OLAP Híbrido) engloba un conjunto de técnicas que tratan de combinar MOLAP y ROLAP de la mejor forma posible. Generalmente puede pre-procesar rápidamente, escala bien, y proporciona una buena función de apoyo.

CUBOS OLAP

Los cubos son elementos claves en OLAP, una tecnología que provee rápido acceso a los datos que se encuentran en el Almacén de datos. Los cubos proveen un mecanismo para buscar datos con rapidez y tiempo de respuesta uniforme independientemente de la cantidad de datos en el cubo o la complejidad del procedimiento de búsqueda.

Los cubos OLAP tienen un número indefinido de dimensiones, razón por la cual también reciben el nombre de hipercubos. Un cubo OLAP contendrá datos de una determinada variable que se desea analizar, proporcionando una vista lógica de los datos provistos por el sistema de información hacia el Datawarehouse, esta vista estará dispuesta según las dimensiones y podrá contener información calculada. El análisis de los datos está basado en las dimensiones del hipercubo, por lo tanto, se trata de un análisis multidimensional.[4]

Dimensiones del CUBO

Las dimensiones del cubo pueden ser creadas para usarse en un cubo individual o en múltiples cubos. Una dimensión creada para un cubo individual, es llamada dimensión privada. Por el contrario, si esta puede ser usada por múltiples cubos, se le llama dimensión compartida. Estas podrán ser usadas dentro de todo cubo, en la base de datos, así se optimiza el tiempo y se evita el andar duplicando dimensiones privadas.

Medidas del CUBO

Las medidas, son datos numéricos de interés primario para los usuarios del cubo. Algunas medidas comunes son ventas en unidades, ventas en pesos, costo de ventas, gastos, conteo de la producción, presupuesto, etc. Estas son usadas por el procedimiento de agregación de los servicios de OLAP y almacenadas para su rápida respuesta a las peticiones de los usuarios.

Se puede crear una medida calculada y calcular miembros de dimensiones, combinando expresiones multidimensionales, fórmulas matemáticas y funciones definidas por el usuario.

Propiedades del CUBO

Se pueden definir propiedades para los miembros de dimensión y usar datos para estas propiedades dentro de un cubo. Por ejemplo, si los miembros de la dimensión producto son sus números de partes, es lo mismo hacer varias propiedades asociadas con este número de parte tales como, el tamaño, color, etc.

Cubos virtuales

Se pueden juntar cubos, dentro de cubos virtuales, muy parecido al proceso de juntar tablas con vistas a las bases de datos relacionales. Un cubo virtual, provee acceso a los datos en los cubos combinados, si la necesidad de construir un nuevo cubo, mientras permite que se mantenga en mejor diseño en cada cubo individual. [4]

Conclusiones.

Teniendo en cuenta las características y conceptos del mundo de la Inteligencia de Negocios se puede resumir de forma general todo su funcionamiento a partir de la Figura 5.

En dicha figura, se observa primero las distintas fuentes de datos que pudieran utilizarse. Después se representa el proceso ETL, este proceso es en el que se definen, de las fuentes heterogéneas, qué campos se van a utilizar, si necesitan algún tipo de modificación o transformación y dónde se quieren ubicar. El tercer elemento de la gráfica representa el repositorio de datos. En este repositorio se encuentran los datos transformados, representados visualmente en modelos multidimensionales, dimensiones y tablas de datos. Existe un proceso entre el repositorio de datos y la interfaz de acceso al usuario, este es el motor de BI que permite habilitar componentes, administrar consultas, monitorear procesos, cálculos, métricas. La interfaz de acceso a usuarios permite interactuar con los datos, representar de forma gráfica los resultados de consultas y los indicadores de gestión que fueron construidos.



Figura 5. Arquitectura de Inteligencia de Negocios

En resumen la Inteligencia de Negocios es una alternativa tecnológica y de administración de negocios de cualquier entidad u organización, que cubre los aspectos del manejo de información para la toma de decisiones, desde su extracción en los sistemas, depuración, transformación, diseño de estructuras de datos o modelos especiales para el almacenamiento de datos, hasta la explotación de la información, mediante herramientas de fácil uso para los usuarios.

REFERENCIAS BIBLIOGRAFICAS

1. Qué es Business Inteligent? (n.d.). Retrieved 5 2011, from Sinnexus.Bussiness Inteligent + Informática Estratégica: http://www.sinnexus.com/business_intelligence
2. DataWarehouse. (n.d.). Retrieved 5 2011, from Sinnexus.Business Intelligent + Informática Estratégica: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx
3. Datamart. (n.d.). Retrieved 5 2011, from Sinnexus.Businness Intelligent +Informática Estratégica: http://www.sinnexus.com/business_intelligence/datamart.aspx
4. Orol, A. M. (n.d.). OLAP y el diseño de cubos| GestioPolis. Retrieved 5 2011, from OLAP y el Diseño de CUBOS: <http://www.gestiopolis.com/canales8/ger/olap-online-analytic-processing.htm>
5. Esquema de la Estrella. (n.d.). Retrieved 5 2011, from ETL-Tools.Info.Business Intelligence - Almacenes de Datos - ET:http://etl-tools.info/es/bi/almacendedatos_esquema-estrella.htm