Tabla

Descripción generada automáticamenteInterfaz de usuario gráfica, Aplicación

Descripción generada automáticamenteDiagrama

Descripción generada automáticamente

Las funciones de un sistema gestor de base de datos multidimensionales

incluyen el modelado inicial de datos en función de las dimensiones seleccionadas, transferencias periódicas en bloque (dado que no es posible transferir datos que están siendo modificados en la base transaccional)

aplicación de precálculo y acumulación en el proceso de transferencia desde la base transaccional a la base multidimensional

rápida respuesta a las consultas intensivas de escenarios de tipo "Qué pasa si",

identificación de patrones para descubrir relaciones insospechadas entre las diferentes dimensiones a través del procesamiento OLAP,

poderoso motor de cálculo y análisis comparativo con potentes funciones estadísticas

pivoteo, tabulación cruzada, comparación entre periodos, profundización mediante funciones de tipo drill down, niveles de resumen por dimensión, etcétera.

La base de datos multidimensional, a diferencia del modelo relacional que utiliza dos dimensiones bajo la forma de tablas con filas y columnas, ofrece una visión multidimensional de los datos que permite la identificación de patrones a partir de una dimensión determinada, aplicando procesamiento OLAP (procesamiento analítico en línea) a diferencia del procesamiento transaccional,

en este tipo de base de datos el objetivo no es el procesamiento transaccional, sino la utilización de datos internos y externos de una organización para analizar posibles escenarios que faciliten la toma de decisiones en lo que se llama inteligencia de negocios.

Las organizaciones entonces tienen dos niveles de procesamiento:

1. **el procesamiento de las operaciones diarias** como compras, ventas, pagos, etcétera, llamado OLTP (Online Transactional Processing o procesamiento transaccional en línea)
2. y, por el otro, el análisis de esos datos para facilitar la planificación estratégica, conocido como OLAP (procesamiento analítico en línea).

Diagrama

Descripción generada automáticamente

El hipercubo OLAP transforma el esquema matricial bidimensional de las tablas relacionales en esquemas multidimensionales lógicos.

Así como en el modelo relacional los datos no están almacenados físicamente como tablas, en el modelo multidimensional no están almacenados como cubos; es un concepto lógico abstracto.

A los datos detallados que se encuentran en la base relacional, el sistema gestor de bases de datos multidimensionales les aplica un algoritmo predeterminado (por ejemplo, acumulación o sumarización), y el resultado es guardado en el cubo unidad que representa las variables dimensiones de acumulación.

Las dimensiones son las áreas temáticas o líneas de negocio definidas como un grupo de uno o más atributos lógicos separados y distintos entre sí, es decir, no comparten atributos. Este esquema ayuda a mostrar múltiples perspectivas de los datos, permitiendo realizar análisis por diferentes dimensiones e incluso cruzando información entre distintas dimensiones.

Por ejemplo, en un hipercubo de tres dimensiones que refleje ventas acumuladas, elegimos un esquema tridimensional por facilidad de representación geométrica, pero podría ser de cualquier número de dimensiones.

Ese espacio dimensional se ha dividido según tres atributos de interés para este negocio:

temporalmente (ventas del primer semestre y ventas del segundo semestre),

por sucursales (sucursal A, B, C o D),

y por artículos (artículo 1, 2, 3 o 4).

Por lo tanto, una unidad del cubo OLAP podría representar las ventas del segundo semestre del artículo 2 en la sucursal C.

El análisis podría hacerse por diferentes dimensiones o planos: por ejemplo, todas las ventas del segundo semestre de todos los artículos en todas las sucursales, todas las ventas del año en la sucursal A, o las ventas anuales del artículo 4 en todas las sucursales. La base multidimensional puede estar formada por varios hipercubos, cada uno con un número variado de dimensiones de acuerdo con los intereses del negocio. Así, podríamos tener otro cubo donde se analicen las ventas de los productos según sus características (color, talle, etc.) para tomar decisiones sobre diversificación de stock o líneas de producción más convenientes.

Diagrama

Descripción generada automáticamente

El análisis OLAP puede realizarse utilizando varias técnicas, siendo las más comunes el *drill down*, *roll up* (desglose y acumulación), y el *slice and dice* (cortes y rotaciones).

En el caso del ***drill down* y el *roll up***, desglosar hacia abajo o hacia arriba es una técnica analítica que permite al usuario navegar entre niveles de datos, pasando de los más resumidos (arriba) a los más detallados (abajo) , aplicando el desglose, o viceversa, sumarizando valores detallados en el caso de la acumulación. Por ejemplo, al ver los datos de ventas de sucursales en la región del Cono Sur, un desglose en la dimensión de región mostraría Argentina, Chile y Brasil, y un análisis más detallado de Argentina mostraría Buenos Aires, Córdoba, etc. Aplicando el criterio inverso, sumaríamos las ventas de Córdoba y Buenos Aires para obtener las ventas de Argentina, y así sucesivamente.

En el caso de cortes y rotaciones *slice and dice*, el usuario navega solicitando vistas de un subconjunto del hipercubo de forma interactiva, a través de la especificación de cortes y rotaciones.

Diagrama

Descripción generada automáticamente

Un almacén de datos o *Data Warehouse* se crea al extraer datos desde una o más bases de datos de aplicaciones transaccionales. Los datos extraídos son transformados para eliminar inconsistencias, resumidos y luego cargados en una base OLAP multidimensional mediante un proceso masivo único que se repite periódicamente, por ejemplo, una vez al día. En la arquitectura del *Data Warehouse*, el primer componente son las fuentes desde las cuales se extraen los datos necesarios para poblar la base de datos multidimensional. Conectados a cada una de estas fuentes se encuentran los extractores, que extraen y transforman los datos. Posteriormente, a través de un proceso integrador temporal, los datos transformados se cargan en la base de datos multidimensional y se generan los metadatos correspondientes para optimizar el acceso a los datos. Es esencial mantener algún tipo de directorio o repositorio de estos datos, donde consten la ubicación de origen, el formato, el algoritmo de transformación aplicado, etc.

Texto, Tabla

Descripción generada automáticamenteDiagrama

Descripción generada automáticamente

Pueden realizarse dos tipos básicos de operaciones algorítmicas en un Data Warehouse:

1-verificación, donde se busca en base a los datos registrados una hipótesis o suposición apriorística,

2- y descubrimiento, donde el objetivo es la identificación de patrones de comportamientos no conocidos o inferidos antes mediante el análisis de datos y la construcción de escenarios posibles.

El **procesamiento estadístico** requiere la definición de una hipótesis que pueda ser respondida por el Data Warehouse, realiza análisis sobre datos elementales de tipo lineal de regresión de variación, etcétera, debe contar con variables conocidas al inicio del análisis y al manipular datos mediante agrupamientos de valores reconoce patrones en series históricas de datos, pivoteo o identificación de valores extremos.

El **procesamiento analítico** evalúa la eficiencia operativa para acciones futuras interrelacionando datos para responder preguntas del tipo Qué pasa así el What if, busca en los datos sin una hipótesis previa construyendo escenarios donde detecta tendencias o patrones recurrentes. Los modelos más sofisticados son sistemas expertos, ellos cuentan con una base de conocimiento que se retroalimenta de sus propios resultados y un motor de inferencia de reglas lógicas, realiza operaciones del tipo análisis de dependencia entre datos con predicción del valor de uno a partir del valor del otro dato, descubrimiento de clases o categorías, construcción de modelos mediante proyecciones de los patrones de los datos analizados y detección de desviaciones que no se ajustan a los estándares.

**Data Mining o minería de datos** lo veremos en detalle en otra presentación.

Aquí se observa un ejemplo de la necesidad de registrar metadatos:

desde tres bases de datos transaccionales se quieren extraer registros basados en el atributo o dimensión de "género"; sin embargo, en cada una de estas bases, el género está registrado en distintos formatos. Por lo que el proceso de transformación selecciona un único formato, ya que no pueden existir formatos diferentes en una base multidimensional. En los metadatos se registra cómo y dónde estaban registrados originalmente cada uno de los atributos antes de la transformación, para poder hacer el camino inverso cuando se quiera aplicar el proceso drill down o desglose para minería de datos

Diagrama

Descripción generada automáticamente

Podemos entender un datamart como un subconjunto de los datos del Data Warehouse, con el objetivo de responder a un determinado análisis, función o necesidad y para una población de usuarios específica.

Al igual que en el Data Warehouse, los datos están estructurados en modelos multidimensionales. Entonces, ¿qué diferencia existe entre un datamart y un Data Warehouse? Su alcance:

el datamart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización, siendo el almacén natural para los datos departamentales.

En cambio, el ámbito del Data Warehouse es la organización en su conjunto, funcionando como el almacén natural para los datos corporativos comunes. Además, los datamarts facilitan el proceso de transferencia hacia los Data Warehouse, ya que los datos extraídos desde los datamarts ya están extraídos y transformados desde las bases transaccionales.

Tabla

Descripción generada automáticamente

para comprender cabalmente el concepto OLAP, a modo de resumen, es conveniente analizar las diferencias con el modelo transaccional. En el modelo transaccional, los datos normalizados están en tablas con filas y columnas, mientras que en el modelo OLAP los datos están desnormalizados y estructurados en hipercubos. El modelo transaccional orienta su estructura a aplicaciones transaccionales, mientras que en OLAP está orientado a dimensiones o patrones de negocio para el análisis estratégico. En el modelo transaccional, los datos no están integrados, mientras que en OLAP deben estarlo para identificar patrones. Además, las bases de datos transaccionales pueden implementarse en diferentes plataformas y modelos estructurales, pero en OLAP se necesita una única plataforma de hardware y un modelo estructural único. Los usuarios del modelo transaccional pueden ingresar, modificar y manipular datos, mientras que en OLAP solo los consultan, ya que estos se transfieren automáticamente desde la base transaccional. Las validaciones en el modelo transaccional se ejecutan a nivel de transacción, mientras que en OLAP se ejecutan antes de la transferencia desde las bases transaccionales. La carga de datos en el modelo transaccional se realiza en cada transacción, mientras que en OLAP se hace en forma masiva mediante procesamiento en lotes. El modelo transaccional almacena datos detallados, mientras que OLAP almacena datos acumulados y resumidos. Finalmente, en el modelo transaccional es prioritario el alto desempeño y la disponibilidad, mientras que en OLAP es prioritaria la flexibilidad y la facilidad de consulta.