BASES DE DATOS MULTIDIMENSIONALES

La base de datos multidimensional a diferencia del modelo relacional que utiliza 2 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 o lap hola qué es el 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 facilitan la toma de decisiones en lo que se llama inteligencia de negocios en las organizaciones entonces tienen 2 niveles de información por un lado el procesamiento de las operaciones diarias compras ventas pagos etcétera llamado o LTP online transaccional processing procesamiento transaccional en línea y por el otro el análisis de esos datos para facilitar la planificación estratégica como ya dijimos hola qué es el procesamiento analítico en línea.

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 cálculo y acumulación el proceso de transferencia desde la base transaccional a la base multidimensional rápida respuesta a las consultas intensivas de escenarios de tipo que pasa así identificación de patrones para descubrir relaciones insospechadas entre las diferentes dimensiones a través del procesamiento olap un poderoso motor de cálculo y análisis comparativo con potentes funciones estadísticas y pivoteo tabulación cruzada comparación entre períodos profundización mediante funciones de tipo drill down niveles de resumen por dimensión etc.

El hipercubo OLAP transforma el esquema 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 las bases de datos multidimensionales les aplica un algoritmo predeterminado simulación por ejemplo humanización y el resultado es guardado en el cubo unidad que represente 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 unos de otros es decir no se comparten atributos y este esquema ayuda a mostrar múltiples perspectivas de los datos permitiendo realizar análisis por diferentes dimensiones incluso cruzando información entre distintas dimensiones sea por ejemplo un hipercubo de 3 dimensiones que refleje ventas acumuladas elegimos un esquema tridimensional por la facilidad de representación geométrica pero podría ser cualquier número de dimensiones ese espacio dimensional se ha dividido según 3 atributos de interés para este negocio por un lado se ha dividido temporalmente en las ventas del primer semestre y ventas del segundo semestre se han tomado en cuenta las ventas en las distintas sucursales sucursal a BC o DI de los distintos artículos artículo 123 o cuatro por lo tanto esta unidad del cubo olap representa las ventas del segundo semestre del artículo 2 en la sucursal se el análisis podría hacerse por diferentes dimensiones o planos por ejemplo podrían analizarse las ventas desde una dimensión temporal todas las ventas de segundo semestre todas las ventas de segundo semestre de todos los artículos en todas las sucursales geográficamente por sucursal todas las ventas del año en la sucursal a producto las ventas anuales del artículo cuatro en todas las sucursales la base multidimensional puede estar formado por varios número variado de dimensiones de acuerdo a los intereses del negocio es decir podríamos tener otro cubo donde se analizan las ventas de los productos según sus características color calle etcétera para tomar decisiones sobre diversificación de stock o líneas de producción más convenientes.

El análisis hola puede hacerse utilizando varias técnicas de las cuales la las más utilizadas son el drill down rolap y la acumulación o el slice andáis cortes y rotaciones veamos cada uno

En el caso del drill hacia abajo o hacia arriba es una técnica analítica específica no hay ante la cual el usuario navega entre niveles de datos que van desde los más resumidos arriba hasta los más detallados abajo estaríamos aplicando acá en este caso el desglose que o viceversa sumar izando valores detallados en este caso acumulación por ejemplo al ver los datos de ventas de sucursales en región cono sur una operación de desglose en la dimensión región mostraría Argentina Chile y Brasil un análisis más detallado de la dimensión Argentina podría mostrar Buenos Aires Córdoba etcétera eh aplicando el criterio inverso estaríamos sumar izando todas las ventas de las regiones de Córdoba más las ventas de las regiones de Buenos Aires con eso tendríamos las ventas de las regiones de Argentina y así sucesivamente.

En el caso de los cortes y rotaciones el proceso de navegación iniciado por el usuario solicita vistas de un subconjunto del hipercubo en forma interactiva a través de la especificación de cortes y rotaciones.

Un almacén de datos o de ihs 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 en un proceso masivo único que se repite periódicamente sólo una vez al día por ejemplo estoy diagrama muestra como primer componente dentro de la arquitectura del data warehouse a las fuentes desde las cuales se extraen los datos necesarios para poblar la base de datos multidimensional conectados a cada una de las fuentes se encuentran los extractores que extraen y transforman los datos de las fuentes posteriormente a través de un proceso integrador temporal estos datos transformados se cargan en la base de datos multidimensional y se generan los metadatos correspondientes a fin de proveer el mejor acceso a los datos es absolutamente necesario mantener alguna forma de directorio o repositorio de estos datos donde conste la ubicación de origen el formato del algoritmo de transformación aplicado, etc.

Vemos aquí un ejemplo de la necesidad de registrar estos metadatos desde bases de datos transaccionales se quieren extraer los registros de acuerdo al atributo o dimensión género pero en cada una de estas bases están asentados con diferente formato por lo que el proceso de transformación selecciona uno no pueden haber diferentes formatos en una única base multidimensional y en los metadatos se registra cómo y dónde estaban registrados cada uno de los atributos antes de la transformación para poder hacer el camino inverso cuando se quiera aplicar el proceso para minería de datos

Realizarse 2 tipos básicos de operaciones algorítmicas en un data warehouse verificación 2 verificación donde se busca confirmar en base a los datos registrados una hipótesis o su posición apriorística 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 delaware house 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 extra el procesamiento analítico evalúa la eficiencia operativa para acciones futuras interrelacionando datos para responder a preguntas del tipo qué pasa si what if busca en los datos sin una hipótesis previa construyendo escenarios donde detecta tendencias o patrones recurrentes en 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. Operaciones de tipo análisis de dependencia de datos con predicción del valor de uno a el otro dato descubrimiento de clases 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.

Podemos entender como un subconjunto de los datos del data con el objetivo de responder a un determinado análisis función o necesidad y para una población de usuarios específica al igual que en el de los datos están estructurados en modelos multidimensionales entonces qué diferencia existe entre un data mart y su alcance el data mart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización es el almacén natural para los dato departmentales en cambio el ámbito de la organización en su conjunto es el almacén natural para los datos corporativos comunes facilitan el proceso de transferencia hacia los mercados ya que los datos extraídos desde ya están extraídos y transformados desde las bases transaccionales.