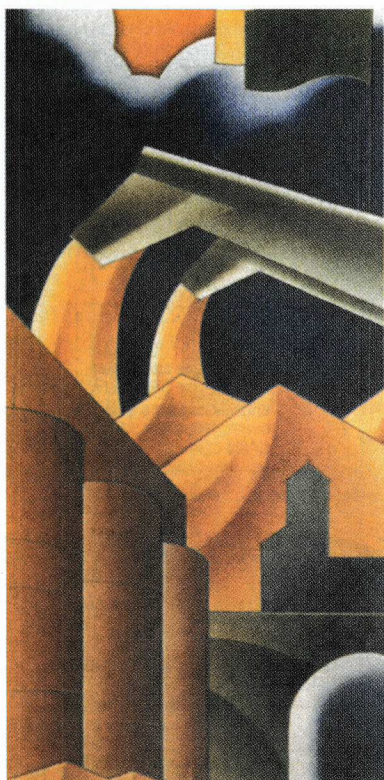


Data Mart

¿Cuánto se ha de cavar?

Data Mart frente a otras tecnologías de extracción de la información



Las bases de datos de cada empresa contienen información que sirve para atender mejor a los clientes, reducir costes y mejorar ventas. Pero realmente hay que realizar una verdadera minería para explotarla. La progresión exponencial de las exigencias del usuario avanzado para explotar esta información, dio origen a varias nuevas tecnologías, cuyas comparaciones y evolución fueron tema de análisis en el número anterior de Windows NT Magazine. Esta vez, enfocaremos de cerca las herramientas Data Mart, utilizando como ejemplo el producto *Hi-Spins* de Software Greenhouse que hoy presentamos.

¿Qué es un Data Mart?

La tecnología Data Warehouse, que se caracteriza por su enorme poder de extracción y presentación de información, tiene sus orígenes en los bancos y las enormes distribuidoras estadounidenses. Esto determina las dimensiones de los proyectos, marcados por un considerable despliegue de procesadores, discos y sobre todo de

consultoría. Utilizar este tipo de proyectos para abordar la problemática de las empresas medias, puede resultar como usar un 747 para ir de Madrid a Alcalá de Henares. Apartando el coste, quien intentó implantar un software escrito para empresas mucho más grandes, sabe que hacer downsizing de software es una de las tareas más ingratas en la informática. Este hecho ha dado origen a los Data Marts como *Hi-Spins*, que adaptan la tecnología Data Warehouse a la problemática de la mediana empresa.

La necesidad de análisis de información existe en la empresa mediana, tanto como en la empresa muy grande,

y sus problemas con la lentitud de las consultas y la necesidad de darles soporte a sus usuarios no son menores. Lo que es diferente es el volumen de información que debe ser sometido al análisis. Por ello y para que los proyectos sean de tamaño razonable y su coste sea aceptable, hay que ajustar algunos conceptos de Data Warehouse.

Basándose en los conceptos de OLAP, Software Greenhouse ha desarrollado su producto, evitando las desventajas de las clásicas soluciones de bases de datos relacionales con índices. Software Greenhouse ha intentado al máximo adaptar la programación, cuyo objetivo es leer en muy corto espacio de tiempo grandes cantidades de registros, a la forma de operar de los discos comúnmente utilizados en los servidores. Es una tecnología más reciente que la de la mayoría de los productos maduros de Data Warehousing actualmente en el mercado que permitirá un fuerte crecimiento de la capacidad de los Data Mart a corto plazo.

La metodología de la puesta en marcha

Los proyectos de *Hi-Spins* duran normalmente de 2 a 4 meses. La primera idea básica es que la base de datos de su empresa es buena. Ha evolucionado de acuerdo con las necesidades de la empresa y por tanto contiene toda la información esencial, aún en el caso que no cumpla con todas las normas. Partiendo de este principio, lo primero que se hace en un proyecto de *Hi-Spins* es examinar la base de

datos y derivar de ella unos modelos multi-dimensionales, que se presentan a los usuarios en forma de maqueta. Una maqueta de Hi-Spins tiene la funcionalidad completa del producto, pero los datos que la llenan, son sólo un modelo aproximado de las futuras dimensiones y variables del modelo.

Este enfoque evita la necesidad de que el usuario tenga que escribir su "carta a los Reyes Magos". ¿Se imaginan un usuario con un lápiz y unas hojas blancas, en las que tenga que indicarnos todas las consultas que querrá hacer en los próximos 5 años? En una maqueta, el usuario puede probar en el teclado el alcance de su futuro sistema de consultas y, por diferencia, determinar qué falta y qué sobra.

Una vez aprobadas las maquetas, comienza la segunda fase del proyecto, en la cual se realiza la conexión con la base de datos operacional. Para la conexión se utilizan los ficheros receptores estándar de Hi-Spins que se colocan en el host. (Este puede ser un AS/400, un servidor NT o Unix con cualquier base de datos comúnmente utilizada o un mainframe.) Los receptores son físicamente idénticos para todas las aplicaciones de Hi-Spins, su contenido, sin embargo, varía de acuerdo con la definición del modelo de datos.

Para realizar la conexión sólo hay que reflejar las columnas de la base de datos de origen, en su sitio correspondiente en los receptores. Para ello, lo más recomendable es utilizar unos programas sencillos en COBOL, C, RPG o cualquier lenguaje común para la instalación. A partir de aquí, el programa de mantenimiento de la base de datos de Hi-Spins se encarga de incorporarlos a la base de datos multi-dimensional.

En la puesta en marcha de Hi-Spins participa el equipo informático de la instalación, conjuntamente con el equipo de Software Greenhouse. Al terminar el proyecto, los informáticos de la instalación están preparados para realizar otras aplicaciones de Hi-Spins

para su empresa. La dependencia del proveedor es prácticamente nula en este sentido.

El Data Mart reside físicamente en los discos del servidor de red y usa sus procesadores. Por ello Hi-Spins no consume los más costosos recursos del ordenador central y su implantación también es económica desde el punto de vista de hardware. Los requerimientos típicos de hardware son: un Pentium 160 o 200, o el nuevo IPCS Pentium para AS/400 y clientes a partir de 486/66 con 8 o preferiblemente 16MB de memoria.

La comodidad del usuario aumenta y con ello su aprecio a la labor del informático

Muchas veces en un solo Data Mart convergen datos de varios ordenadores, así como datos de fuentes externas como la información del mercado, precios de materia prima (provenientes de la bolsa de valores) etc. También están disponibles interfases con los más importantes productos de software de aplicaciones, presentes en el mercado.

La metadata de Hi-Spins define los modelos multi-dimensionales y el contenido de los receptores. No se usa, sin embargo, para la conexión con la base de datos original. Dado el hecho que estas bases de datos no son un ejemplo de una demo, sino el resultado real de una evolución compleja de una instalación, el uso de metadata para definir la extracción, adquiere una complejidad excesiva. La paradoja resultante es que para evitar la programación, tenemos que aprender otro nuevo lenguaje de algo - metadata, CASE - que no se quiere llamar programación. A muchos

usuarios de CASE esto le sonará familiar.

No obstante, para los que definitivamente prefieren evitar la programación, existen los productos con tecnología mirroring, que pueden ser utilizados para la conexión. Mirroring también puede facilitar la conexión cuando se requiere que el Data Mart se actualice en tiempo real, aunque no es indispensable para ello.

Hi-Spins y los usuarios

Los beneficios principales que aporta Hi-Spins a los usuarios son la velocidad de las consultas y la posibilidad de que sean ellos mismos, quienes sin conocer la base de datos ni entender los modelos multi-dimensionales, especifiquen las consultas requeridas. Con ello se superan los dos problemas principales de los productos "tipo query". Un usuario que quiere saber, ¿Cuáles de sus clientes, pertenecientes al Grupo IFA, no han comprado durante los últimos dos meses los productos de la marca Heno de Pravia? y puede formular su pregunta en 20 segundos y obtener la respuesta en 5 segundos más, está listo para modificar todo el concepto que se haya formado de la informática. Y si su empresa tiene un concepto claro de la productividad del trabajo administrativo, no dudará un segundo sobre la utilidad de Hi-Spins.

Un usuario con conocimiento regular de Windows, sólo necesita de 3 a 6 horas de entrenamiento para aprender a sacar todo el beneficio de Hi-Spins. La principal diferencia con la tecnología EIS es que un Data Mart no necesita ningún tipo de mantenimiento para formular nuevas consultas. Por esta razón, algunas instalaciones que ya tenían EIS han instalado Hi-Spins para aquellos usuarios, cuyas consultas son mucho más variadas que las de la alta dirección. Los usuarios creativos con capacidad de improvisación son los que más se benefician de Hi-Spins. Para las consultas repetitivas, Hi-Spins tiene los catálogos de consultas predefinidas.

La interfase de Hi-Spins con el usuario permite tomar todas las decisiones de navegación de una forma muy sencilla. Es un producto hecho para trabajar, no para hacer demos, por lo que intenta evitar aquellos "efectos especiales" que consumen muchos recursos de hardware y a la larga no siempre facilitan el entendimiento de una consulta. Se apoya cada vez más en una interfase bidireccional con hojas electrónicas. Esto significa que no sólo puede convertir una consulta en hoja electrónica, sino también *usar hojas electrónicas como entrada de datos* (con una validación completa y la comprobación de autorizaciones necesarias). Hi-Spins es capaz de reconocer determinadas hojas e identificar la posición que corresponde a cada celda en el modelo multi-dimensional. Este enfoque *significa una gran simplificación para algunas tareas administrativas, tales como cálculo y seguimiento de presupuestos*.

Funcionalmente, Hi-Spins cumple todos los requerimientos de Data Mart, presentando además algunas características sumamente importantes, que muchos otros productos no ofrecen. Entre ellas se puede nombrar la definición de consultas "línea por línea", jerarquías múltiples (Clientes por Tipo, por Cadena, por Vendedor, por Zona etc., a la vez), selecciones y omisiones múltiples, prueba de presencia o ausencia de registros ("no han comprado"), selecciones por expresión con múltiples variables, o utilización de hojas electrónicas como input.

¿Si existiese un generador de tiempo disponible, lo compraría?

En la mayoría de instalaciones que usan productos para extraer información de la base de datos, los usuarios consumen considerables cantidades de tiempo de los informáticos. El proyecto de conexión de Hi-Spins con la base de datos de origen, resuelve el problema de una sola vez y el mantenimiento requerido es mínimo. Hi-

Un pequeño ejercicio de Economía

Intentemos calcular los beneficios económicos que puede producir **Hi-Spins** en el primer año de su explotación, en una instalación con 10 usuarios, que dedican aproximadamente 20% de su tiempo al trabajo con la información. Consideremos que un usuario de query que ahora dedica a este tipo de trabajo una hora y media diaria, con **Hi-Spins** hará lo mismo en 30 minutos.

Concepto	Horas/ Día	Días/ Año	Ptas/ Hora	Número Personas	Total Ptas.
Ahorro del tiempo de los usuarios	1	220	4.000	10	8.800.000
Ahorro en el tiempo de soporte	4	220	4.000	1	3.520.000
Hardware (ahorro estimado)					3.500.000
Coste del proyecto					-4.000.000
Total					11.820.000

Sólo una observación: Esta tabla no incluye los beneficios más importantes - ¡los intangibles producidos por la substancial mejora en el análisis de la información!

Spins reduce el soporte para las consultas de los usuarios a la mínima expresión. Su modelo multi-dimensional puede reflejar toda la información contenida en la base de datos operacional, y proporciona a los usuarios una facilidad de acceso sin precedentes. *La utilidad de la informática y la comodidad de uso para el usuario final aumenta enormemente y con ello también su aprecio de la labor del informático.*

Hi-Spins está preparado para el año 2000. Soporta no sólo el año de 4 dígitos, sino también múltiples variables monetarias con factores de conversión flexibles entre ellas. Cada programa de output que puede ser sustituido por consultas de Hi-Spins, significa un programa menos para someter a la conversión.

Hi-Spins proporciona al informático APIs y servidores OLE para realizar, a distintos niveles, la interfase con el producto. Esto hace posible, por un lado, presentar las consultas bajo cualquier formato desarrollado en Visual Basic, Delphi u otros lenguajes de interfases gráficas. Por otro lado, desde cualquier aplicación, ya sea Excel, Word o aplicaciones propias de la instala-

ción, pueden invocarse consultas de Hi-Spins, bajo el formato de hojas electrónicas o el formato propio de Hi-Spins.

El diálogo con los usuarios

En las líneas anteriores, hemos intentado profundizar en la materia Data Mart, usando el ejemplo del producto Hi-Spins de Software Greenhouse. El hecho que para nosotros este producto se identifique con todas las virtudes de la tecnología Data Mart, no es casual. Lo desarrollamos con determinada filosofía, porque creemos que debe hacerse de esta forma. Pero Hi-Spins, más que ningún otro producto anterior nuestro, es resultado de un constante diálogo con los informáticos y usuarios de nuestras instalaciones. Por esto le sugerimos llamarnos, para que le demos la lista completa de las instalaciones de Hi-Spins. Nuestros usuarios son los que mejor les podrán informar, hasta que punto nuestro enfoque ha dado resultados positivos en su instalación. ▲

Para más información, contacte con Jorge Gros, de Software Greenhouse, S.A.
(93) 417.83.50; jgros@swgreenhouse.com