



Sybase IQ

Servidor analítico con arquitectura basada en columnas



Sybase IQ

Descripción



Tener acceso a toda la información de que dispone su organización, con el fin de analizarla no es hoy una opción, es una necesidad de cualquier empresa que quiera competir en su mercado. Sólo si disponemos de una solución que nos facilite la toma de decisiones críticas para nuestro negocio en el momento adecuado tendremos la capacidad de descubrir cambios, adaptarnos y obtener ventajas diferenciales frente a la competencia.

No es suficiente disponer de una solución que nos permita el análisis de la información, debemos ser exigentes, necesitamos información concreta y en el momento adecuado, sin demoras, lo que nos obliga a demandar rapidez y flexibilidad en el momento de obtener datos, y no a cualquier coste.

El ahorro de costes, debido a su rápida implantación y el bajo nivel de recursos hardware necesarios, su flexibilidad a la hora de interrogar los datos y su novedoso sistema de análisis capaz de ofrecer respuestas a sus inquietudes de negocio en segundos, hacen de Sybase IQ la mejor plataforma para el almacenamiento y análisis de información del mercado.

Sybase ha patentado una tecnología exclusiva para la carga, almacenamiento y análisis de información que proporciona ventajas únicas para el desarrollo e implantación de soluciones de Business Intelligence. Esta tecnología, orientada al almacenamiento por columnas, es el núcleo del Gestor de Base de Datos analítico llamado Sybase IQ.

Sybase IQ, dentro de la arquitectura de Sybase para Data Warehouse, está específicamente diseñado para el almacenamiento en columnas de los datos, de forma que el acceso a la información se realiza de forma inmediata e independiente de la consulta que se realice. Evitando de esta forma el obtener una vez tras otra el mismo informe, el cual responde a una pregunta predefinida de difícil modificación por las consecuencias que éste hecho tendría sobre la velocidad de respuesta del sistema.

A continuación se detallan los conceptos básicos sobre los que se basa esta tecnología. Es importante hacer constar que todo lo que se explica es totalmente transparente tanto para el administrador del sistema como para el usuario del mismo. Para ellos Sybase IQ se ve como un gestor de base de datos relacional estándar, exactamente igual que si se estuviera trabajando con ORACLE, DB2, SQL Server, etc. Únicamente se indica a efectos informativos para que el lector comprenda el porqué técnico de las ventajas que proporciona Sybase IQ.

Tecnología al servicio del negocio

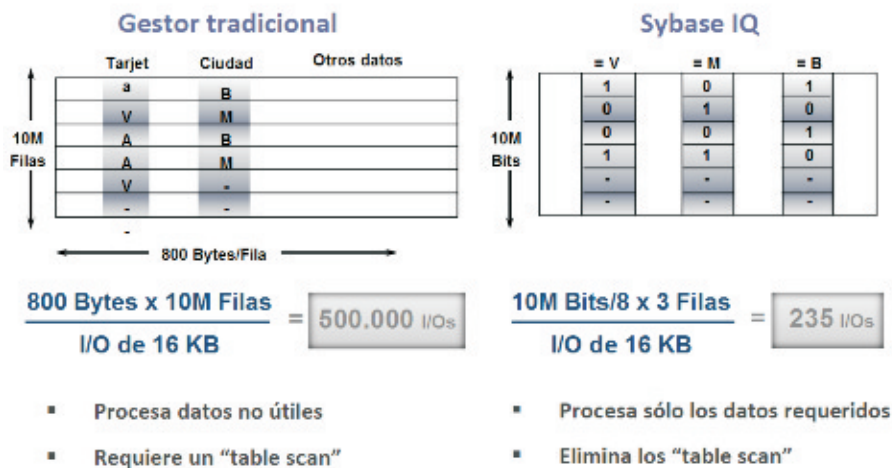
Normalmente un Gestor de Base de Datos Relacional no está diseñado para dar respuesta a queries "ad hoc" (es decir queries no definidas a priori). El gestor busca un índice por donde ejecutar la query, en la mayoría de los casos no existirá tal índice, con lo que el gestor acabará haciendo una lectura secuencial completa de toda la información, con el consiguiente aumento en la carga de trabajo, en particular cuando la query involucra a varias tablas en un "Join".

Esto, que no es un problema en un entorno transaccional, ya que la mayoría de los accesos a la Base de Datos están predefinidas se convierte en el principal cuello de botella en sistemas de análisis en donde el usuario se conecta y requiere información de forma interactiva. La filosofía de Sybase IQ está totalmente orientada a este tipo de queries.

Todas las columnas se encuentran indexadas (los propios datos son los índices) de forma que el gestor no realiza las búsquedas leyendo las filas (un tanto por ciento muy elevado de los datos de la fila no se utilizan en la query) sino “entrando” directamente por las columnas (solamente aquellas que interesan en la query).

A continuación se muestra un ejemplo que aclara este modo de funcionamiento.

“¿Cuántos clientes con tarjeta Visa existen en Madrid y Barcelona?”



Los datos se almacenan en el sistema por columnas en formato “Bit Wise”. Este formato es una ampliación a los formatos tradicionales de Bit Maps, con las adaptaciones necesarias para dar respuestas a los principales problemas de este tipo de soluciones.

Habitualmente, los formatos Bit Maps utilizados por el resto de gestores únicamente son válidos para columnas de baja cardinalidad (menos de 100 valores diferentes). Para este caso Sybase IQ usa los índices llamados LF (que permiten hasta aproximadamente 1000 valores diferentes).

En el caso de las columnas con mayor cardinalidad, Sybase IQ usa la tecnología Bit Wise consistente básicamente en calcular el valor binario y posteriormente separar este valor por columnas. De esta forma Sybase IQ, a diferencia del resto de gestores, permite almacenar en este formato cualquier tipo de información.

Este método de almacenamiento, no sólo nos permite una consulta más rápida de la información, sino también su almacenamiento óptimo haciendo uso de menores recursos de disco y máquina.

¿Sybase IQ por qué ?



Información al instante

Velocidad en la resolución de consultas. Sybase IQ proporciona tiempos de respuesta miles de veces más rápidos que un Gestor de Base de Datos Relacional. Esto quiere decir que consultas que en un sistema tradicional pueden llegar a durar horas, sobre Sybase IQ tardarían segundos.

Este es quizás el punto más crítico en los sistemas de análisis. Debido al gran volumen de datos que se almacenan y al tipo de 'queries' que involucran a casi la totalidad de dichos datos, el usuario ve como el tiempo que el sistema utiliza para responder a sus consultas es demasiado elevado, llevándole a no utilizarlo o seguir dejando al departamento de informática que sean ellos los que 'resuelvan' estas consultas en modo 'Batch'.

Sybase IQ resuelve completamente dicho problema. El usuario obtendrá siempre sus respuestas en cuestión de segundos, lo que le llevará a utilizar cada vez más el sistema para obtener más información. Esto redundará en una amortización más rápida de la implantación y una mayor satisfacción.

Velocidad de carga

Carga de Datos. En los SGBDR, independientemente del fabricante del gestor, los tiempos de carga de una base de datos siempre se dividen en 3: Carga de los datos, Creación de índices y Cálculo de sumalizaciones y agregaciones.

En Sybase IQ este carga se realiza en un único paso. Esto es debido a que se reducen los conceptos de índices o Sumalizaciones. Lo único que existe y que se carga directamente son los datos, de esta forma se pueden alcanzar tiempos de carga del orden de 10 veces inferior a los necesarios por un Gestor de Base de Datos Relacional. Este ratio se corresponde al tiempo total de carga, es decir no existen otros tiempos suplementarios.

Esta disminución en los tiempos de carga puede suponer la diferencia entre que el Sistema sea operativo o no. Sybase IQ va a permitir que el usuario final disponga de la información actualizada de forma inmediata con lo que se aumenta el valor real para el usuario de dicha información.

Menor espacio en disco

Disminución de los costes asociados a la ocupación de los datos. En un Gestor de Base de Datos Relacional usado para entornos de Data Warehouse el volumen real de la información no son solamente los datos originales, sino que se debe sumar la ocupación de los índices y de las sumalizaciones con lo que habitualmente la ocupación real crece aproximadamente al triple. En el caso de Sybase IQ debido a la compresión de los datos la ocupación real es incluso inferior a la de los datos originales.

De forma general y aproximada se puede afirmar que la necesidad de volumen de almacenamiento de datos en Sybase IQ es entre 3 y 6 veces inferior al requerido por un Data Warehouse construido con un Gestor de Base de Datos Relacional.

Por lo tanto, para un Data Warehouse de 6 TB de almacenamiento total, Sybase IQ únicamente necesitaría 1 TB, con lo que estaríamos ahorrando el coste necesario para los 5 TB restantes. Solamente con este ahorro ya se amortizaría directamente el coste de la puesta en marcha del Sistema.

Hardware

Disminución de los costes asociados a los recursos Hardware. Una 'creencia' muy extendida es que la puesta en marcha de un sistema de Business Intelligence lleva asociado una inversión muy alta en recursos hardware (procesadores, discos, memoria, etc.).

Esto es real en el caso de usar un Gestor de Base de Datos analítico con arquitectura tradicional, ya que su propia arquitectura basada en el paralelismo, requiere de una infraestructura Hardware muy potente para soportarla.

En el caso de Sybase IQ, no se requiere tanta potencia de máquina, ya que su particular diseño y funcionamiento, suple el uso de recursos Hardware.

De esta forma, existe un importante ahorro de costes asociados a la compra de infraestructura Hardware. Se han llegado a implementar sistemas de Business Intelligence con Sybase IQ soportando Terabytes de información y cientos de usuarios con máquinas Windows o Linux de bajo coste.

Escalabilidad

Escalabilidad ilimitada con la arquitectura Multiplex Sybase IQ. Se ha demostrado con numerosas pruebas realizadas que dicha arquitectura garantiza un comportamiento totalmente lineal (tiempos de respuesta estables) en las condiciones de no saturación de recursos físicos de la máquina que lo soporta.

Esta forma de funcionamiento del gestor demuestra su versatilidad para adecuar a cualquier situación su estabilidad media de tiempos de respuesta.

Administración

Disminución de los costes asociados a los recursos de administración. El gran esfuerzo de administración que requieren este tipo de sistemas (debido en gran parte al gran volumen de datos y a la complejidad de las consultas que soporta) es uno de los puntos críticos para su viabilidad y evolución.

Al no utilizar el concepto tradicional de los gestores de bases de datos de índices y tablas auxiliares o de agregaciones, Sybase IQ necesita menos administración y 'tuning'. El concepto se asemeja a 'crea el modelo físico de datos, carga los datos y olvídate'. Esta 'casi nula' administración facilita su mantenimiento, incluso para personas sin conocimientos previos en este tipo de soluciones. Con esto se consigue una reducción en los costes que supone tener un gran número de administradores cualificados en entornos de bases de datos analíticas.

Concurrencia

Soporte a elevada concurrencia de usuarios. Uno de los principales 'cuellos de botella' a los que se enfrentan los Gestores de Bases de Datos tradicionales como por ejemplo los relacionales o Multidimensionales, es el aumento de los tiempos de respuestas, cuando existen múltiples usuarios accediendo concurrentemente a consultar la información.

Sybase IQ, gracias a su tecnología exclusiva orientada a columnas puede dar soporte a miles de usuarios concurrentes realizando consultas complejas sobre grandes volúmenes de información. Esta ventaja permite dar un mejor uso a un mayor número de usuarios, sin tener que invertir más dinero en la compra de nuevos equipos o en la ampliación de los ya existentes.

Arquitectura estándar

Interface transparente. Sybase IQ se comporta de cara al administrador del sistema o al usuario final exactamente igual que un Gestor de Base de Datos Relacional, siendo su forma de acceso con SQL estándar, drivers ODBC/JDBC y certificado con todas las herramientas de análisis del mercado. De esta forma es posible la puesta en marcha del sistema sin ningún tipo de conocimiento adicional de nuevas tecnologías. Igualmente, con esta arquitectura, es posible el cambio de un Data Warehouse ya en funcionamiento sobre cualquier SGBDR sin ninguna modificación en la arquitectura global del sistema, en las herramientas de análisis usadas o en los desarrollos ya realizados. Con lo que sin ningún coste adicional conseguiremos evitar los problemas que podremos estar 'sufriendo' con dicho sistema (Malos tiempos de respuesta, elevado coste de administración, poco uso, gran utilización de recursos hardware, etc.)

Sybase IQ

Arquitectura



La arquitectura Multiplex del gestor Sybase IQ nos permite aumentar la configuración del sistema incrementando el número de máquinas (no sólo de procesadores). Con esto podemos garantizar que conseguimos los mismos resultados que estamos obteniendo actualmente pero multiplicados (o divididos según lo que estemos midiendo, tiempo de respuesta o carga de trabajo por tiempo) por el número de máquinas que añadamos.

Esta arquitectura garantiza que el crecimiento será lineal, ya que no estamos añadiendo más recursos de un tipo a una máquina (CPU, memoria, canal de entrada salida) sino que lo que estamos haciendo es aumentar por el doble o el triple el sistema que tenemos actualmente y que ya hemos comprobado como se comporta.

Este crecimiento es inmediato y transparente. No se necesita realizar una reorganización del sistema como ocurre en otro tipo de arquitecturas como las basadas en gestores de bases de datos en paralelo.

Técnicamente esta solución se basa en la posibilidad de definir tantos nodos lógicos o físicos como deseemos sobre una única Base de Datos física, de forma que los usuarios, independientemente de a que nodo se conecten 'ven' siempre los mismos datos.

Esto permite poder configurar cada nodo de una forma diferente para cada tipo de usuario que se conecte así como garantizar la escalabilidad lineal a medida que se van conectando nuevos usuarios.

Igualmente se puede utilizar esta tecnología para mantener soluciones de Alta Disponibilidad en los que cuando 'cae' uno de los nodos todas sus conexiones son migradas al resto de nodos.

