

Dining Experience

Se refiere al proceso de extraer, transformar y cargar datos de varias fuentes en un repositorio central para informes y análisis. Esto suele implicar el uso de software y herramientas especializadas para limpiar, estandarizar y organizar los datos, y luego almacenarlos en un formato que está optimizado para la consulta y el informe. El objetivo final es hacer que sea fácil para los analistas y los usuarios de negocios acceder y comprender los datos, para que puedan tomar decisiones mejores y mejorar el rendimiento de sus organizaciones.

Two Powerful Ideas

Las dos ideas poderosas para construir almacenes de datos exitosos: separar sistemas y construir estrellas y cubos. El autor enfatiza la importancia de la comprensibilidad del usuario final y la velocidad de ejecución de consultas como restricciones no negociables en el diseño del almacén de datos. También reconocen la complejidad del espacio de diseño del almacén de datos y la necesidad de descomponer el problema en partes manejables y utilizar técnicas predecibles, reutilizables y robustas para navegar por esta complejidad.

Separate Your Systems

El primer paso en el diseño de un almacén de datos en una empresa compleja, que es separar los sistemas lógica, física y administrativamente. El autor sugiere pensar en el proyecto como cuatro sistemas distintos: sistemas de procesamiento de transacciones de origen (fuente), sistemas de área de preparación del almacén de datos, sistemas de presentación del almacén de datos y herramientas analíticas de alto rendimiento opcionales. El administrador del almacén de datos solo debe ser responsable del área de preparación de datos y del sistema de presentación. El área de preparación de datos es donde se limpian, se conforman y combinan los datos de producción antes de ser entregados a los sistemas de presentación del almacén de datos. El sistema de presentación está construido para mejorar la experiencia de consulta e informes para los usuarios finales. Las herramientas analíticas de alto rendimiento opcionales consumen los datos del almacén de datos en lotes para tareas como minería de datos, pronóstico, puntaje y asignación.

Symmetrical Stars and Cubes

el uso de esquemas de estrella relacionales y cubos OLAP multidimensionales como las estructuras de datos dominantes en la mayoría de las áreas de presentación, ya que se ha demostrado que son fácilmente comprensibles para los usuarios finales, y la simplicidad de estas estructuras de datos permite desarrollar algoritmos poderosos para una rápida ejecución de consultas. La simetría tanto del esquema estrella como del cubo OLAP también permite interfaces de usuario previsibles, escenarios administrativos previsibles en el área de preparación y respuestas de implementación previsibles cuando se dispone de nuevos tipos de datos. Los esquemas de estrella son los más apropiados para conjuntos de datos grandes con muchas medidas numéricas o miembros, mientras que los cubos OLAP son los mejores para conjuntos de datos más pequeños donde las herramientas analíticas pueden realizar comparaciones y cálculos complejos. El autor recomienda alimentar los

datos en una estructura de esquema de estrella antes de transformarlo en un cubo OLAP, y utilizar sistemas híbridos de esquema estrella-OLAP para suavizar los objetivos de perforación de cubos OLAP más pequeños bajo una única interfaz de usuario.

The Big Payoff

los beneficios de utilizar esquemas estrella simétricos y cubos OLAP en el sistema de presentación del almacén de datos. La principal ventaja es el conjunto previsible de puntos comunes para conectar datos de toda la empresa. El autor sugiere que, en la próxima columna, se discutirán técnicas para conformar las definiciones de las dimensiones y los hechos en todas las fuentes de datos dispares de una gran empresa. Las dimensiones conformadas y los hechos formarán la base para una arquitectura de bus de almacén de datos, un conjunto de puntos de conexión estándar que proporcionan poder al almacén de datos, similar a la barra bus en una estación de energía o el bus en una computadora que proporciona datos a todos los periféricos.

Conclusión

El texto discute la implementación de dos ideas poderosas en el proceso de diseño de almacenes de datos. La primera idea es separar los sistemas lógica, física y administrativamente en cuatro tipos distintos, con el administrador de almacén de datos siendo responsable solo de dos de ellos. Esta separación permite una separación de las responsabilidades incompatibles de la preparación de datos y las consultas de usuarios finales. La segunda idea es poblar el área de presentación del almacén de datos con esquemas de estrella y cubos OLAP, ya que estas estructuras son comprensibles, rápidas y pueden manejar consultas ad-hoc. Estas dos ideas han abordado efectivamente muchas de las restricciones y realidades complejas en el diseño de almacenes de datos, incluyendo la comprensibilidad, la velocidad de consulta, los costos, los riesgos de centralización inapropiada, el desarrollo incremental, el manejo del cambio continuo y el papel de los data marts.