

Resumen

Two Powerful Ideas / Data Warehouse Dining Experience

Ralph Kimball en su publicación “Two Powerful Ideas”, los almacenes de datos más exitosos se enfocan o entornan en dos ámbitos. Primero es la separación de sistemas de negocio y Segundo la construcción de estrellas y cubos donde podemos transformar, calcular y presentar dicha información.

Esta identificación de sistemas de trabajo oportuno, apoya al diseñador de un almacén de datos, a la forma correcta de trabajo, entendiendo objetivos claves como la capacidad de comprensión del usuario final y la velocidad de ejecución de consultas. Como bien menciona es importante separar sistemas, para entender el entorno de trabajo, sabiendo las fronteras o limitantes que no deben cruzarse entre los mismos.

Es de suma importancia en un almacén de datos separar los sistemas tanto de manera lógica, física y administrativa. De estos podemos determinar los siguientes sistemas.

1. Sistemas de procesamiento de transacciones de producción (fuente).
2. Sistemas de área de almacenamiento de datos.
3. Sistemas de presentación de almacenamiento de datos, que incluyen redactores de informes y herramientas de consulta basadas en web y cliente/servidor.
4. Herramientas analíticas opcionales de alta gama que admiten la extracción de datos, la previsión, la puntuación o la asignación.

Dado esta identificación, contemplamos en el primer sistema de trabajo es el responsable del almacén de datos, donde los datos de producción de muchas fuentes se introducen, limpian, conforman, combinan y, en última instancia, se envían a los sistemas de presentación del almacén de datos. Se ha escrito mucho sobre los pasos cruciales de extracción-transformación-carga (ETL) en el área de preparación, pero alejándose de este detalle, el principal requisito para el área de preparación es que está **“fuera del alcance de todos los clientes finales del almacén de datos”**.

Gracias a este proceso de transformación, Margy Ross, en su escrito “Data Warehouse Dining Experience”, nos ejemplifica de una manera paralela de trabajo sobre un restaurante y un almacén de datos. Esto con lleva que antes de poder estar en producción se debe planificar de la manera más conveniente y eficaz para elaborar el diseño y cada componente del espacio de trabajo.

Si nos basamos en el siguiente dibujo de un restaurante contemporáneo que cada noche debe atender a sus comensales con calidad, eficiencia y eficacia; esto con el fin de tener a un cliente fijo que pueda visitar el restaurante continuamente.

Diseño y Construcción de Data Warehouses

Tarea No. 1



Primero, el diseño debe ser muy eficiente, esto con el fin de facilitar el trabajo adentro del área de transformación de un plato. Ofrecer una calidad constante de la cocina del restaurante es el segundo objetivo y también debe ser de alta integridad.

Debemos enfocarnos que nuestros clientes o usuarios finales, si se encuentran satisfecho con nuestro trabajo presentado, si es fácil de comprender para ellos; querrán seguir nutriendo cada vez más esta información que los ayude a identificar mejoras en su negocio. Ya que esto para ellos es una inversión cuantiosa el tener un almacén de datos y debe saber representar esta necesidad de negocio a la simplicidad de la información y presentación de la misma.

En esto, si nos basamos en el diseño de un restaurante que es un factor importante para atraer al cliente mientras este espera el plato, todo este diseño debe ir acorde de diseño de restaurante y menú elaborado.



El área de preparación debe diseñarse y diseñarse mucho antes de que se extraigan datos de la fuente. Al igual que la cocina, el área de preparación está diseñada para garantizar el rendimiento. Debe transformar los datos de origen sin procesar en el modelo de destino de manera eficiente, minimizando el movimiento innecesario si es posible.

Retomando sobre esta analogía de un restaurante y un almacén de datos. En términos de almacenamiento de datos, al excluir a todos los clientes del almacenamiento de datos del área de almacenamiento de datos, evitamos:

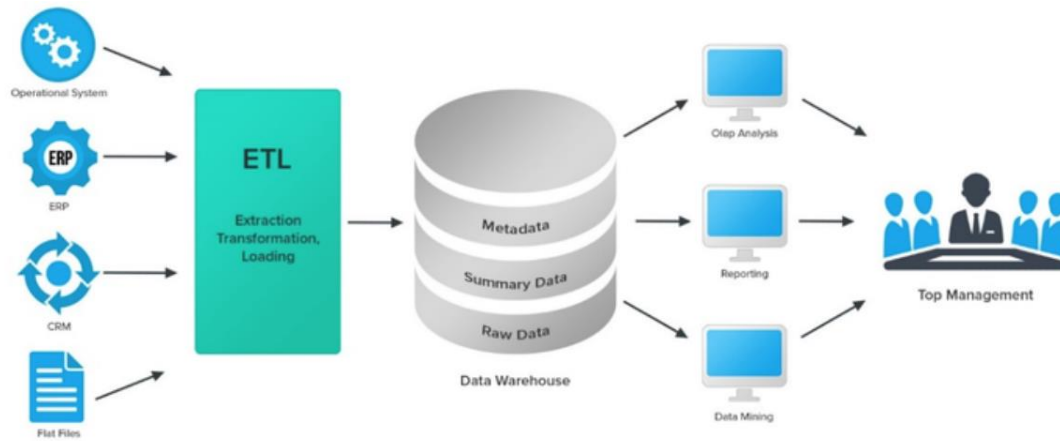
- Garantizar niveles de servicio de tiempo de actividad para consultas o informes
- Hacer cumplir la seguridad a nivel de cliente
- Crear índices y agregaciones que mejoran el rendimiento para el rendimiento de las consultas
- Manejo de conflictos lógicos y físicos entre los pasos de consulta y limpieza de datos
- Garantizar la coherencia entre fuentes de datos independientes y asincrónicas.

Las dos estructuras de datos dominantes en el área de preparación de datos son el archivo plano y el esquema de entidad / relación, que se extraen o derivan directamente de los sistemas de producción. Casi todo el procesamiento en el área de preparación es clasificación o procesamiento secuencial simple. El segundo sistema principal bajo el control específico del almacén de datos es el sistema de presentación

Las estructuras de datos dominantes en el área de presentación son el esquema en estrella relacional y el cubo de datos de procesamiento analítico en línea (OLAP). El procesamiento en el área de presentación debe responder a una avalancha de consultas grandes y pequeñas que llegan desde todos los ángulos posibles sobre los datos.

la minería de datos como disciplina es una tarea interpretativa compleja que involucra una colección completa de técnicas analíticas poderosas, muchas de las cuales la comunidad de usuarios finales no comprende o no confía fácilmente. La minería de datos adecuada requiere un experto profesional en minería de datos que esté equipado para usar las

herramientas de manera efectiva y representar los resultados de la minería de datos a la comunidad.



La entrega oportuna, efectividad y eficiencia para las organizaciones son puntos de atención y dependerá del estudio estratégico, la administración asertiva, la aplicación de Business Process Management (BPM) la apropiada BI para llegar al cumplimiento exitoso (Shahzad & Zdravkovic, 2011).

La mayoría de las áreas de presentación actuales están dominadas por esquemas en estrella relacionales y cubos de datos OLAP multidimensionales. Estas estructuras de datos han demostrado que son las que los usuarios finales pueden comprender.

La simplicidad de los esquemas en estrella y los cubos OLAP ha permitido a los diseñadores de software inteligentes centrarse en algoritmos muy potentes para un rápido rendimiento de consultas.

La simetría tanto del esquema en estrella como del cubo OLAP también contribuye a:

- Interfaces de usuario predecibles que pueden "aprender" qué hacer en el momento de la consulta
- Escenarios administrativos predecibles en el área de preparación porque todas las estructuras de datos tienen la misma apariencia familiar
- Respuestas de implementación predecibles cada vez que se ponen a disposición nuevos tipos de datos.

En casi todos los entornos de cubos OLAP, se recomienda que origine los datos en una estructura de esquema en estrella y luego utilice asistentes para transformar los datos en el cubo OLAP. Los sistemas híbridos de esquema en estrella-OLAP permiten que conjuntos de datos muy grandes en formato de esquema en estrella sean objetivos de desglose sin problemas de cubos de datos OLAP más pequeños, todo bajo una única interfaz de usuario.

Diseño y Construcción de Data Warehouses
Tarea No. 1

La gran recompensa final por construir el sistema de presentación en el almacén de datos alrededor de esquemas en estrella simétricos y cubos OLAP es el conjunto predecible de puntos en común para vincular datos de toda la empresa.