

UNIVERSIDAD AUTÓNOMA “GABRIEL RENÉ MORENO”

FACULTAD DE INGENIERIA EN CIENCIAS DE LA
COMPUTACION Y TELECOMUNICACIONES



METODOLOGIAS PARA EL DESARROLLO DE DATA WAREHOUSE

NOMBRE: Jose Carlos Girón Solano

REGISTRO: 216161924

MATERIA: Sistemas para el soporte a la toma de decisiones

AÑO: 2024

Kimball

Iniciado por Ralph Kimball, el modelo de datos de Kimball sigue un enfoque de abajo hacia arriba para almacenamiento de datos (DW) diseño de arquitectura en el que los mercados de datos se forman primero en función de los requisitos comerciales.

A continuación, se evalúan las fuentes de datos primarias y se utiliza una herramienta de extracción, transformación y carga (ETL) para obtener diferentes tipos de formatos de datos de varias fuentes y cargarlos en un área de ensayo del servidor de base de datos relacional. Una vez que los datos se cargan en el área de preparación en el almacén de datos, la siguiente fase incluye la carga de datos en un modelo de almacén de datos dimensional que está desnormalizado por naturaleza. Este modelo divide los datos en la tabla de hechos, que son datos transaccionales numéricos, o la tabla de dimensiones, que es la información de referencia que respalda los hechos. El esquema en estrella es el elemento fundamental del modelo de almacén de datos dimensional. En este esquema en estrella, una tabla de hechos está limitada por varias dimensiones. El modelado dimensional de Kimball permite a los usuarios construir varios esquemas en estrella para satisfacer diversas necesidades de informes. La ventaja del esquema en estrella es que las consultas de tablas dimensionales pequeñas se ejecutan instantáneamente.

Para integrar datos, la arquitectura de Kimball DW sugiere la idea de dimensiones de datos conformadas. Existe como una tabla de dimensión básica compartida en diferentes tablas de hechos (como cliente y producto) dentro de un almacén de datos o como las mismas tablas de dimensión en varios data marts de Kimball. Esto garantiza que un solo elemento de datos se utilice de manera similar en todos los hechos.

Una herramienta de diseño importante en la metodología de almacenamiento de datos de Ralph Kimball es la matriz de bus empresarial o la arquitectura de bus de Kimball que registra verticalmente los hechos y registra horizontalmente las dimensiones conformadas. La matriz de Kimball, que forma parte de la arquitectura del bus, muestra cómo se construyen los esquemas en estrella. Los equipos de gestión empresarial lo utilizan como entrada para priorizar qué fila de la matriz de Kimball se debe implementar primero.

El enfoque de Kimball también se basa en hechos conformados, es decir, mercados de datos que se implementan por separado junto con una arquitectura robusta.

Ventajas del método Kimball

Algunos de los principales beneficios de la metodología Kimball incluyen:

- El modelado dimensional de Kimball es tan rápido de construir ya que no implica normalización, lo que significa una ejecución rápida de la fase inicial del almacenamiento de datos diseño de procesos.
- Una ventaja del esquema en estrella es que la mayoría de los operadores de datos pueden comprenderlo fácilmente debido a su estructura desnormalizada, que simplifica las consultas y el análisis.
- La huella del sistema de almacenamiento de datos es trivial porque se centra en áreas y procesos comerciales individuales en lugar de en toda la empresa. Por lo tanto, ocupa menos espacio en la base de datos, lo que simplifica la administración del sistema.
- Permite la recuperación rápida de datos del almacén de datos, ya que los datos se segregan en tablas de hechos y dimensiones. Por ejemplo, la tabla de hechos y dimensiones para la industria de seguros incluiría transacciones de pólizas y transacciones de reclamaciones.
- Un equipo más pequeño de diseñadores y planificadores es suficiente para la gestión del almacén de datos porque los sistemas de origen de datos son estables y el almacén de datos está orientado a procesos. Además, la optimización de consultas es sencilla, predecible y controlable.
- Estructura dimensional conformada para marco de calidad de datos. El enfoque de Kimball también se conoce como el enfoque de estilo de vida dimensional empresarial porque permite que las herramientas de inteligencia empresarial profundicen en varios esquemas en estrella y genere información confiable.

Enfoque de Kimball para el ciclo de vida del almacén de datos (Fuente: Kimball Group)

Desventajas del método Kimball

Algunos de los inconvenientes del enfoque de diseño de Kimball incluyen:

- Los datos no están completamente integrados antes de la presentación de informes; la idea de una "fuente única de verdad se pierde".
- Pueden ocurrir irregularidades cuando los datos se actualizan en la arquitectura Kimball

DW. Esto se debe a que, en el almacén de datos de técnicas de desnormalización, se agregan datos redundantes a las tablas de la base de datos.

- En la arquitectura Kimball DW, pueden ocurrir problemas de rendimiento debido a la adición de columnas en la tabla de hechos, ya que estas tablas son bastante detalladas. La adición de nuevas columnas puede expandir las dimensiones de la tabla de hechos, afectando su rendimiento. Además, el modelo de almacén de datos dimensional se vuelve difícil de modificar con cualquier cambio en las necesidades comerciales.
- Como el modelo de Kimball está orientado a los procesos comerciales, en lugar de centrarse en la empresa en su conjunto, no puede manejar todos los requisitos de informes de BI.
- El proceso de incorporar grandes cantidades de datos heredados en el almacén de datos es complejo.

Cree su almacén de datos desde cero

Diseñe, pruebe, lance e implemente el almacén de datos desde cero y automatice los procesos para brindar información rápidamente sin escribir una sola línea de código.

El metodo Inmon

Bill Inmon, el padre del almacenamiento de datos, ideó el concepto de desarrollar un almacén de datos que comienza a diseñar el modelo de datos del almacén de datos corporativos, que identifica las principales áreas temáticas y entidades con las que trabaja la empresa, como clientes, producto, proveedor, y así. La definición de Bill Inmon de un almacén de datos es que es una "colección de datos orientada al sujeto, no volátil, integrada y variable en el tiempo en apoyo de las decisiones de la gerencia".

Luego, el modelo crea un modelo lógico completo para cada entidad primaria. Por ejemplo, se construye un modelo lógico para productos con todos los atributos asociados con esa entidad.

Este modelo lógico podría incluir diez entidades diversas bajo producto, incluidos todos los detalles, como impulsores comerciales, aspectos, relaciones, dependencias y afiliaciones.

El Enfoque de diseño Inmon utiliza la forma normalizada para construir la estructura de la entidad, evitando la redundancia de datos tanto como sea posible. Esto da como resultado una identificación clara de los requisitos comerciales y la prevención de irregularidades en la actualización de datos. Además, la ventaja de este enfoque de arriba hacia abajo en el diseño de

bases de datos es que es robusto a los cambios comerciales y contiene una perspectiva dimensional de los datos en el mercado de datos.

A continuación, se construye el modelo físico, que sigue la estructura normalizada. Este modelo de Inmon crea una única fuente de verdad para todo el negocio. La carga de datos se vuelve menos compleja debido a la estructura normalizada del modelo. Sin embargo, utilizar esta disposición para realizar consultas es un desafío, ya que incluye numerosas tablas y enlaces.

Esta metodología de data warehouse de Inmon propone la construcción de data marts por separado para cada división, como finanzas, marketing, ventas, etc. Todos los datos que ingresan al data warehouse están integrados. El almacén de datos actúa como una única fuente de datos para varios data marts a fin de garantizar la integridad y coherencia en toda la empresa.

Ventajas del método Inmon

El enfoque de diseño de Inmon ofrece los siguientes beneficios:

- El almacén de datos actúa como una fuente de verdad unificada para todo el negocio, donde todos los datos están integrados.
- Este enfoque tiene una redundancia de datos muy baja. Por lo tanto, hay menos posibilidad de irregularidades en la actualización de datos, lo que hace que el proceso de almacenamiento de datos ETL sea más sencillo y menos susceptible a fallas.
- Simplifica los procesos comerciales, ya que el modelo lógico representa objetos comerciales detallados.
- Este enfoque ofrece una mayor flexibilidad, ya que es más fácil actualizar el almacén de datos en caso de que haya algún cambio en los requisitos comerciales o en los datos de origen.
- Puede manejar diversos requisitos de informes en toda la empresa.

Desventajas del método Inmon

Los posibles inconvenientes de este enfoque son los siguientes:

- La complejidad aumenta a medida que se agregan varias tablas al modelo de datos con el tiempo.
- Se requieren recursos capacitados en el modelado de datos de almacenamiento de datos,

que pueden ser costosos y difíciles de encontrar.

- La configuración preliminar y la entrega requieren mucho tiempo.
- Se requiere una operación ETL adicional, ya que los data marts se crean después de la creación del almacén de datos.
- Este enfoque requiere que los expertos administren un almacén de datos de manera efectiva.