

METODOLOGÍAS PARA CONSTRUIR UN DATA WAREHOUSE

METODOLOGÍA HEFESTO

HEFESTO es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, comparación de metodologías existentes, experiencias propias en procesos de confección de almacenes de datos. Cabe destacar que HEFESTO está en continua evolución, y se han tenido en cuenta, como gran valor agregado, todos los feedbacks que han aportado quienes han utilizado esta metodología en diversos países y con diversos fines.

La idea principal, es comprender cada paso que se realizará, para no caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo, ni por qué.

La construcción e implementación de un DW puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que para algunas fases en particular, las acciones que se han de realizar serán muy diferentes. Lo que se debe tener muy en cuenta, es no entrar en la utilización de metodologías que requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del DW y motivar a los usuarios.

La metodología HEFESTO, puede ser embebida en cualquier ciclo de vida que cumpla con la condición antes declarada.

Con el fin de que se llegue a una total comprensión de cada paso o etapa, se acompañará con la implementación en una empresa real, para demostrar los resultados que se deben obtener y ejemplificar cada concepto.

Esta metodología cuenta con las siguientes características:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos de los usuarios, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tome decisiones respecto al comportamiento y funciones del DW.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.

- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- Se aplica tanto para Data Warehouse como para Data Mart.

METODOLOGÍA SAS

El sistema de análisis social (SAS) es presentado por Jacques Chevalier como un conjunto de herramientas participativas de investigación-acción fáciles de usar para abordar aspectos sociales del desarrollo con múltiples actores de grupos involucrados que puedan incidir en procesos o ser afectados por estos.

La propuesta de investigación del SAS busca la integración de los momentos de la investigación con la práctica concurrente de los proyectos, combinando equilibradamente la recolección de información con la interpretación de la misma y la aplicación práctica de sus conclusiones.

En su propuesta de integrar momentos y elementos, el SAS también se propone constituir diversos niveles de análisis, varias disciplinas y diferentes sistemas de aprendizaje. En ese sentido es una propuesta interdisciplinaria, que busca adaptarse tanto a los niveles populares, como académicos, pasando por una amplia gama de actores.

En término de los actores, destaca en su identificación amplia (incluyendo sus vínculos más allá de lo local o comunidades locales), en sus perfiles, posiciones, involucramiento y empoderamiento.

Incluye elementos derivados de la investigación-acción-participativa, la economía política, la economía formal, la gestión administrativa, la antropología social y la psicología.

El SAS plantea que añade a la literatura existente nuevas formas de conducir un análisis social participativo y el manejo de procesos en los cuales se emplea una planificación continua y flexible frente a situaciones que no pueden ser previstas.

En esencia el SAS propone romper con las separaciones entre diferentes procesos que aparecen divididos en la práctica y buscar uniones ideales. Por ejemplo:

- Investigación y acción (proyectos)

- Conocimiento científico y conocimiento común
- Planificación y evaluación (o monitoreo)
- Objetos de la investigación y sujetos de la acción
- Análisis cuantitativo con cualitativo
- Diferentes expresiones de las ciencias
- Investigación y aprendizaje
- Local y global (glocal)

Al final se propone como un conjunto de técnicas combinables y aplicables según la situación de que se trate. Es en este sentido una propuesta de flexibilidad de aplicación.

METODOLOGÍA 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. Para garantizar la integridad y

la coherencia en toda la empresa, el almacén de datos actúa como una única fuente de datos para varios data marts.

Ventajas del método Inmon

El enfoque de diseño de Inmon ofrece los siguientes beneficios:

- El almacén de datos actúa como una fuente unificada de verdad para toda la empresa, donde se integran todos los datos.
- 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.

METODOLOGÍA KIMBALL

Ralph Kimball inició el enfoque del almacén de datos de Kimball, donde 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. En el Enfoque de abajo hacia arriba de Kimball, un Después de que los datos se cargan en el área de preparación del almacén de datos, la siguiente fase incluye cargar 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 tabla de dimensiones, que es la información de referencia que respalda los hechos.

El esquema de estrella es el elemento fundamental de la dimensión. En este esquema en estrella, una tabla de hechos está limitada por varias dimensiones. Se pueden construir varios esquemas en estrella dentro de un modelo dimensional de Kimball 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 que se comparte 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 debe implementarse 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 de diseño de Kimball DW incluyen:

- Los beneficios del modelado dimensional de Kimball son que es rápido de construir ya que no hay normalización involucrada, 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 bastante 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.

Desventajas del método Kimball

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

- En el diseño ETL de Kimball, 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 está orientado a los procesos comerciales, en lugar de centrarse en la empresa en su conjunto, el diseño de Kimball 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.