

Diseño basado en Data Warehouse

Los Almacenes de Datos (AD) o Data Warehouse son una colección de datos históricos que sirven de apoyo a la toma de decisiones para mejorar un proceso.

Las empresas utilizan los datos acumulados durante años, los datos de las diferentes actividades se almacenan y consolidan en una base de datos central denominada Data Warehouse.

Un Data warehouse es una colección de datos no volátiles, que varían en el tiempo, que están orientados a un tema determinado y que se utilizan para tomar decisiones organizacionales

CARACTERÍSTICAS DE UN AD

Orientado a temas. La información se clasifica en base a los intereses de la organización

Integrado. Recibe este término cuando sus datos provienen de diversas fuentes

No-volátiles. La información es útil para el análisis y toma de decisiones solo cuando es estable

Variables en el tiempo. Debido al gran volumen de información los resultados de una consulta tardarán en originarse por eso se dice que el AD es de tiempo variable.

IMPORTANCIA DE UN AD

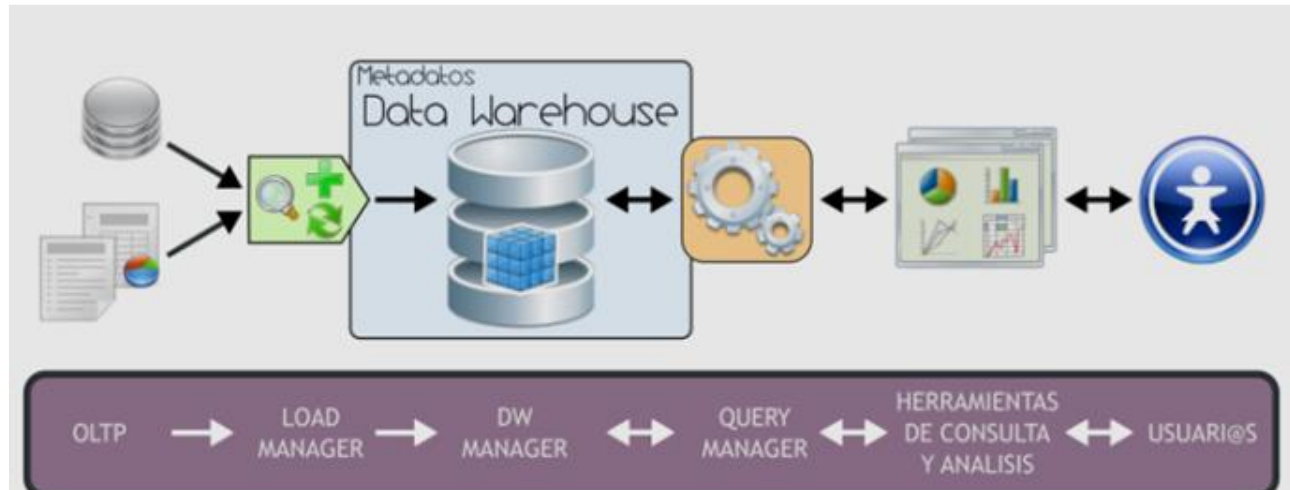
- Mejora la entrega de información
- Facilita el proceso de toma de decisiones
- Mejor integración y automatización en procesos empresariales

METODOLOGÍA HEFESTO

ARQUITECTURA

- Los datos se extraen de aplicaciones, bases de datos, archivos, entre otros.
- Los datos pueden tener distinta arquitectura o formato. Así que tienen que ser integrados, transformados y limpiados.
- Posteriormente se cargan en el AD con una estructura conceptual de cubos multidimensionales.
- Los usuarios acceden a los cubos de información a través de distintas herramientas.

COMPONENTES



- **OLTP (On Line Transaction Processing).** Toda información que genera la organización y fuentes externas diariamente
- **LOAD MANAGER.** Los ETL se encargan de extraer datos desde OLTP para manipular, integrar, transformar y cargar resultados. Estos sistemas son necesarios en un AD
- **DW MANAGER.** Transformar e integrar los datos fuente y de almacenamiento intermedio a un modelo adecuado para la toma de decisiones.
- **QUERY MANAGER.** Recibe consultas del usuario, las aplica a la estructura de datos correspondiente y devuelve los resultados obtenidos.
- **HERRAMIENTAS Y CONSULTAS DE DATOS.** Sistemas que permiten al usuario realizar exploración de datos del AD.
- **USUARIOS.** Aquellos que se encargan de tomar decisiones y planificar actividades del negocio

PASOS Y APLICACIÓN

1. Análisis de Requerimientos
 - a. Identificar preguntas
 - b. Identificar indicadores y perspectivas de análisis
 - c. Modelo conceptual
2. Análisis de los OLTP
 - a. Determinación de Indicadores
 - b. Establecer correspondencia
 - c. Nivel de granularidad
 - d. Modelo conceptual ampliado
3. Modelo lógico del Almacén de Datos
 - a. Tipo del modelo lógico del Almacén de Datos
 - b. Tabla de dimensiones
 - c. Tabla de hechos
 - d. Uniones
4. Procesos ETL

1. Se debe iniciar con la definición del alcance, para evitar grandes problemas en sus fases, se recomienda usar la técnica de descomposición de tareas, y para su representación normalmente se usa la forma tipo organigrama.
2. Se debe redactar un plan inicial que contemple la visión del negocio junto con las metas y la justificación del proyecto.
3. Se deberán definir roles dependiendo de las actividades de cada usuario.
4. Se debe realizar un documento que describa las características de la empresa que contemple datos que la identifiquen, objetivos, políticas, estrategias, definición de los procesos y la relación de las metas con el AD.
5. Como siguiente paso se realiza la identificación de preguntas mediante cuestionarios, entrevistas u observaciones sobre los objetivos del proyecto, con la finalidad de identificar los indicadores y perspectivas de las cuales partirá el análisis de diseño.
6. Una vez que se han establecido las preguntas claves, se debe proceder a su descomposición para descubrir los indicadores que representan lo que se desea analizar concretamente y las perspectivas que se refieren a los objetos mediante los cuales se quiere examinar los indicadores, realizando un detalle de estos.
7. A partir de los indicadores y perspectivas obtenidas en el paso anterior se debe construir un modelo conceptual.
8. Se realiza la representación de los casos de uso que representan los requerimientos del AD.
9. La fuente de datos del AD es el modelo entidad-relación representado en diagramas UML. Al transformar el modelo ER a un diagrama de clases se transforma en el esquema conceptual de la fuente de datos (SCS). En el diseño conceptual del AD según DWEP se realiza los diagramas: conceptual de datos (SCS), mapeo de datos (DMS), almacén de datos (DWCS) y el esquema conceptual del cliente (CCS). El esquema conceptual de la base de datos se realiza en tres niveles que consisten en la realización de un detalle de cómo se encuentra integrada.
10. Este nivel finaliza con la redacción de un acta de aprobación del cliente respecto a la definición de indicadores, fuente de datos y el modelo conceptual.

DWEP

Metodología proceso de Ingeniería para el desarrollo de Almacenes de Datos (DWEP)

FASES

- **Fase de inicio.** Descripción general del proyecto, riesgos críticos, y establecer funcionalidad básica del software
- **Fase de elaboración.** Arquitectura sólida y base lógica para la construcción del software
- **Fase de construcción.** Desarrollar un producto listo para la operación inicial en base a la arquitectura antes planteada
- **Fase de transición.** Se busca implantar completamente el producto en su entorno de operación

FLUJO DE TRABAJO

- **Requerimientos.** Durante este flujo de trabajo, los usuarios especifican las medidas y agregaciones más interesantes, el análisis dimensional, consultas usadas para la generación de reportes periódicos y frecuencia de la actualización de los datos. El PU sugiere el uso de casos de uso [9] [4]. Esto ayuda a comprender el sistema y obtener los requisitos y funciones para la solución. Además establece como deben ser las interacciones del sistema.
- **Análisis.** Tiene como objetivo mejorar la estructura y los requisitos obtenidos en la etapa de requerimientos. En esta etapa se documentan los sistemas operacionales preexistentes que alimentarán el AD. El PU propone el uso del diagrama de clase [4] [9].
- **Diseño.** Al final de este flujo de trabajo, está definida la estructura del AD. El principal resultado de este flujo de trabajo es el modelo conceptual del AD.
- **Implementación.** Durante este flujo de trabajo, el AD es construido y se empiezan a recibir datos de los sistemas operaciones, se afina para un funcionamiento optimizado, entre otras tareas.
- **Pruebas.** El objetivo de este flujo de trabajo es verificar que la aplicación funcione correctamente, realizar las pruebas y analizando los resultados de cada prueba.
- **Mantenimiento.** Un AD es un sistema que se retroalimenta constantemente. El objetivo de este flujo de trabajo es definir la actualización y carga de los procesos necesarios para mantener el AD.
- **Revisiones post desarrollo.** Esto no es un flujo de trabajo de las actividades de desarrollo, sino un proceso de revisión para la mejora de proyectos a futuro. Si hacemos un seguimiento del tiempo y esfuerzo invertido en cada fase es útil en la estimación de tiempo y de las necesidades para generar los requisitos para desarrollos futuros.