

DATA WHERE HOUSE

¿QUE ES?

Es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

Un data Waterhouse funciona **como un repertorio central**. La información proviene de una o varias fuentes de datos, como sistemas transaccionales u otras bases de datos relacionales.

Los datos pueden ser estructurados, semiestructurados o no estructurados. Una vez que están integrados en el Waterhouse, son tratados y transformados. Los usuarios pueden luego acceder a ellos con la ayuda de herramientas de Business Intelligence, de clientes [SQL](#) o de tablas.

Diferentes tipos de Data Warehouse

Data Warehouse de Empresas

Son depósitos de datos centralizados que permiten orientar las decisiones de la empresa.

Data Stores Operacionales

Los datos se actualizan en tiempo real, lo que los hace muy útiles para actividades cotidianas como el registro de informes y de empleados.

Data mart

Es una subcategoría de Data Waterhouse. Está concebida para empresas de sectores de la venta o las finanzas. Los datos pueden ser recolectados desde diversas fuentes.

Estados de un Data Warehouse

Sin conexión: los datos son copiados desde un sistema operacional hacia otro servidor. La carga, el tratamiento y los informes de datos no afectan el rendimiento del sistema operativo.

En línea: los datos se actualizan regularmente desde la base de datos operacional. Los datos se actualizan cada vez que una transacción se realiza en la base de datos operacional. Por ejemplo, podemos mencionar un sistema de reservación de trenes o aviones.

Estructuras de un Data Waterhouse

- **Con una estructura básica**, sistemas operativos y archivos planos proporcionan datos en bruto que se almacenan junto con metadatos. Los usuarios finales pueden acceder a ellos para su análisis, generación de informes y minería.
- **Al añadir un área de ensayo** que se puede colocar entre las fuentes de datos y el almacén, ésta proporciona un lugar donde los datos se pueden limpiar antes de entrar en el almacén. Es posible personalizar la arquitectura del almacén para diferentes grupos dentro de la organización.
- **Se puede hacer agregando data marts**, que son sistemas diseñados para una línea de negocio en particular. Se pueden tener data marts separados para ventas, inventario y compras, por ejemplo, y los usuarios finales pueden acceder a datos de uno o de todos los data marts del departamento.

Diferentes componentes de un Data Warehouse

load manager: permite todas las operaciones de extracción y carga de datos hacia el depósito. También se encarga de la transformación de datos.

Warehouse manager

se ocupa de las operaciones relacionadas con la gestión de datos dentro del depósito.

Administrador de búsquedas

responsable de las operaciones relativas a la gestión de búsquedas de usuarios para orientarlos hacia las tablas apropiadas.

Ventajas de los Data Warehouse

1. Resultan útiles para permitir a las empresas acceder de forma rápida y fácil a los datos provenientes de diversas fuentes de forma centralizada.
2. acceder a informaciones coherentes y actualizadas sobre todas las actividades de la empresa
3. Permiten generar informes y efectuar búsquedas con los datos.
4. permite reducir el tiempo necesario para el análisis de datos y la producción de informes y facilita estas tareas
5. los usuarios pueden analizar las tendencias sobre diferentes periodos temporales para realizar predicciones futuras.

Inconvenientes de los Data Warehouse

1. no es la mejor solución para los datos no estructurados.
2. la creación e implementación de depósitos de datos llevan tiempo y requieren en general mucho trabajo.
3. es difícil realizar cambios en el tipo de datos, los esquemas de fuentes de datos, los index y las búsquedas.
4. La utilización de una plataforma como esta puede resultar demasiado compleja para un usuario promedio.

Características

Orientada al negocio

Sólo ingresan datos relevantes para el análisis y toma de decisiones

Ejemplo: direcciones de habitación, códigos postales, direcciones de correo, entre otros.
Pero son de interés variables como, tipo de cliente, ubicación geográfica, edad, etc.

Integrada

Todos los datos provenientes de fuentes heterogéneas se consolidan para garantizar su calidad y limpieza.

Los principales orígenes de datos son:

Según el tipo de usuario.

- Operacional: Diariamente produce gran cantidad de datos, pero por sí mismos son de poca relevancia para el análisis requerido. Por ejemplo, las ventas de productos.
- Medio: Genera datos con implicación a corto y mediano plazo, basados en los datos operacionales. Un buen ejemplo de este concepto es la generación de inventarios.
- Gerencial: Emplea datos resultantes del proceso de integración y transformación. A su vez, genera nueva información. Se refiere, básicamente, al usuario del Data Warehouse.

Según el área o el departamento de la organización

-
- Áreas: Cada una tiene responsabilidades bien definidas. Producen datos propios que son compartidos con las demás áreas.
- Subdivisiones: Suelen ser geográficas. Aportan datos de localización, que deben ser incorporados junto con los demás.

Según la fuente

-
- Internas: Generan datos propios, provenientes de las actividades diarias de la empresa.
- Externas: Complementan los datos internos, por ejemplo censos y estadísticas.

Variante en el tiempo

Permite el acceso a diferentes versiones de una misma situación, ya que los datos actuales son almacenados junto con los históricos, en los ejemplos de data warehouse.

No volátil

manipulados únicamente cuando se introducen y cuando se consultan.

Riesgos de aplicación

Requiere una gran inversión por parte de la organización. Los beneficios de su implementación no se ven a corto plazo, sino a mediano y largo plazo.

La manipulación de datos atenta contra la manipulación de los datos sensibles.

Cualidades

Maneja datos en volumen, consecuencia de la acumulación de datos históricos, actuales y agregados, provenientes de diversas fuentes.

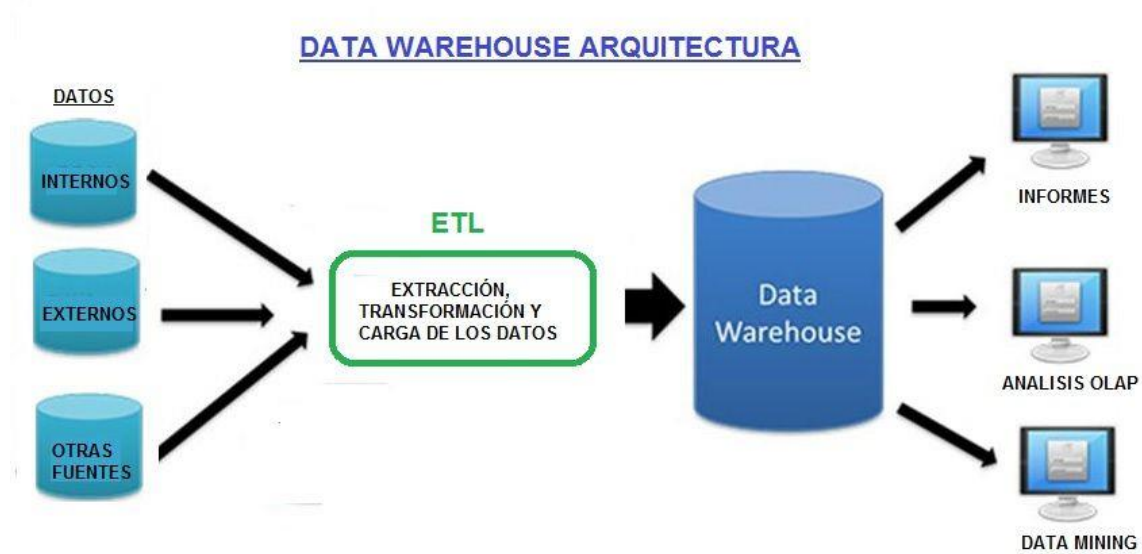
Costos de aplicación

Un data Warehouse conlleva costos de construcción, operación y de sostén. El costo de construcción implica los costos de recursos humanos, tiempo y tecnología, mientras que el de operación y mantenimiento, contempla los costos de evolución, crecimiento y los producidos por los cambios en el origen de los datos.

Impacto en las personas

La aplicación de un Data Warehouse siempre genera expectativas en los usuarios, quienes necesariamente tendrán que adquirir nuevas destrezas. El éxito de este tipo de data depende del uso activo y retroalimentación por parte de los usuarios.

ARQUITECTURA



Este sistema implica una serie de interacciones entre sus componentes. Al respecto y a manera de resumen.

su funcionamiento puede describirse de la siguiente forma:

- Los datos son tomados de diversas fuentes, como servicios web, archivos y otras bases de datos, tanto internas como externas.
- Una vez extraídos los datos, estos son integrados, transformados y limpiado, para posteriormente ser cargados en el Data Warehouse.
- Con la finalidad de generar información táctica y estratégica, se obtienen informes y análisis provenientes de la carga de los datos.
- Finalmente, los usuarios pueden consultar y explorar los reportes y análisis generados.

Elementos que puede utilizar para su funcionamiento

Fuentes del Data Warehouse

son el resultado de la actividad diaria de la empresa, en cuyo caso se denominan fuentes internas. Cuando los datos son tomados de, por ejemplo, servidores web, éstas son consideradas fuentes externas.

Extracción, transformación y carga

es el proceso que comprende todas las tareas que se realizan desde que se obtienen los datos hasta que se cargan en el Data Warehouse

Extracción

Incluye técnicas enfocadas en obtener, desde diversas fuentes, sólo los datos relevantes y mantenerlos en un almacenamiento interno.

Transformación

técnicas encargadas de hacer compatibles los diferentes formatos, así como de filtrar y clasificar los datos, y relacionar fuentes.

Carga

- La carga inicial se refiere a la primera carga de datos que recibe el Data Warehouse. Generalmente, consume muchísimo tiempo debido a la gran cantidad de registros pertenecientes a largos periodos de tiempo.
- La actualización periódica se refiere a la inserción de pequeños volúmenes de datos. Su objetivo es añadir a los ejemplos de data warehouse sólo aquellos datos que se generan de la última actualización. Depende de las necesidades y requerimientos del usuario.

Reportes

son herramientas gráficas que le permiten al usuario obtener informes detallados sobre la información de su empresa.

OLAP

Es el componente más poderoso del Data Warehouse, ya que contiene el motor de consultas multidimensionales especializadas del sistema.

Data Mining

una herramienta estadística, mediante la cual se pueden hacer predicciones. Se trata de inferir comportamientos, sin que hayan reglas preestablecidas.

Bases de datos tradicionales vs Data Warehouse

- En las bases de datos tradicionales la información está organizada para que sea recuperada y actualizada fácilmente. Un Data Warehouse está organizado y orientado hacia el usuario final, el cual únicamente puede realizar consultas.
- Las bases de datos transaccionales se ocupan del procesamiento diario de los datos. El Data Warehouse trabaja con datos históricos, es decir, correspondientes a largos períodos de tiempo.
- Las bases de datos tradicionales son accedidas en varias oportunidades durante una jornada de trabajo. En un Data Warehouse las lecturas y consultas son mínimas, pues se accede esporádicamente.
- El volumen de datos que administra un Data Warehouse es mucho mayor que el administrado en las bases de datos tradicionales.
- La estructura de las bases transaccionales es estable. La estructura de un Data Warehouse varía de acuerdo con su propia evolución y utilización.

Recomendaciones para ver..

[\(49\) ¿Qué es un Data Warehouse? | Business Intelligence ⚡ - YouTube](#)

[\(49\) Data Warehousing - An Overview - YouTube](#)