

Nombre : Kasandra Mamani Rodríguez

BODEGA DE DATOS (DATAWAREHOUSE)

1. ¿Qué es un Data Warehouse ?

Es un dispositivo 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.

Normalmente, un data warehouse se aloja en un servidor corporativo o cada vez más, en la nube. Los datos de diferentes aplicaciones de procesamiento de transacciones Online (OLTP) y otras fuentes se extraen selectivamente para su uso por aplicaciones analíticas y de consultas por usuarios.

2. ¿En qué consiste el Data Warehouse ?

Un Data Warehouse supone una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla. De esta manera, es posible analizar dicha información desde diferentes puntos de vistas y a gran velocidad. Se trata de unos de los componentes más importantes de la inteligencia empresarial en el entorno actual en el que operan las empresas.

3. Características del Data Warehouse

- **Orientación a temas :** Solo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas, con el fin de facilitar su acceso y entendimiento por parte de los usuarios.
- **Variante en el tiempo :** Los datos muestran el estado de la actividad del negocio en ese mismo momento. Por otra parte, la información almacenada en el Data Warehouse se utiliza para realizar análisis de tendencias, debido a la posibilidad de realizar comparaciones entre los datos en el almacenados.
- **No volátil :** La información no se puede modificar ni eliminar una vez almacenado el dato. Solo podrá ser usado como lectura o consulta en un futuro.
- **Integrado :** Los datos almacenados deben estar formados por una estructura consistente. Esta información puede estructurarse en distintos niveles de detalles para adaptarse de mejor manera a las distintas necesidades de los usuarios.

- **Metadatos**: Cuenta con metadatos es decir, datos sobre datos. Los metadatos permiten simplificar y automatizar la obtención de la información de una manera precisa, ayudando a localizar los datos que nos interesan. El empleo de metadatos es análogo al empleo de índices.

4. Ventajas del Data Warehouse

- Herramienta de toma de decisiones en cualquier área, basándose en la información integrada y global de la empresa.
- Permite la aplicación de técnicas de análisis para encontrar relaciones ocultas entre los data warehouse.
- Facilita dentro de la empresa la implantación de sistemas de gestión integral con relación al cliente.

5. Estructuras de un Data Warehouse

La arquitectura de un data warehouse puede ser dividida en tres estructuras simplificadas: básica, básica con un área de ensayo y básica con un área de ensayo y data marts.

- **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 minerías.
- **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.

5. Metodologías de Data Warehouse

1. Multidimensional (Bottom up/Ascendente)

La metodología multidimensional o metodología de Ralph Kimball, mantiene un diseño ascendente. Por lo que los Data Marts son los primeros en crearse y después se integran al Data Warehouse, donde se busca que el almacenamiento de datos de los usuarios se execute de la forma más rápida posible.

Con esta metodología, al crear primero los Data Marts se proporcionan capacidades analíticas de informes para procesos específicos de negocio y funcionales.

Ventajas

- No requiere un equipo muy grande de desarrolladores y arquitectos de datos para mantener el Data Warehouse (menor Costo)
- Brinda buena funcionalidad y seguimientos de las métricas o KPI's orienta los Data Marts a informes en cuanto a procesos de departamento o de negocios.
- La optimización de las consultas es sencilla, predecible y controlable.

2. Relacional (Top Down / Descendente)

La metodología relacional de Bill Inmon muestra un diseño descendente, donde se construye primero el Data Warehouse y posteriormente el Data Marts. Ubicando el Data Warehouse en el centro de la información corporativa lo que asegura un marco lógico en los datos.

Ventajas

- Proporciona una única versión de la verdad, al ser el único origen de datos para los Data Marts.
- Tiene una mayor facilidad de comprensión de los procesos empresariales para los usuarios.
- Resulta más fácil y menos propenso al fracaso el proceso de ETL
- Mayor flexibilidad a cambios de necesidades analíticas, de negocio

3. Data Vault

Es una metodología híbrida creada por Don Linsted, es utilizada principalmente cuando las empresas tienen un crecimiento exponencial constante de datos por lo que presentan problemas de rediseño y mantenimiento.

Ventajas

- Diseñado especializado para almacenar registros (hace el proceso de registro de datos sea más sencillo).
- Automatiza fácilmente los procesos ETL
- Fácil rastreo y auditoria de datos
- Permite varios sistemas de origen y relaciones con cambios frecuentes.