



SIO: SISTEMA INFORMÁTICO ORGANIZACIONAL

Materia

W. BUR

Tutor

Data Warehouse

Un data warehouse 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.

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 de otras fuentes se extraen selectivamente para su uso por aplicaciones analíticas y de consultas por usuarios.

Data Warehouse es una arquitectura de almacenamiento de datos que permite a los ejecutivos de negocios organizar, comprender y utilizar sus datos para tomar decisiones estratégicas

Un poco de historia

Cuando aparecieron por primera vez los data warehouses a finales de los 80 (desarrollado por Paul Murphy y Barry Devlin de IBM), su objetivo era que los datos pasaran de los sistemas operativos a los sistemas de apoyo a las decisiones (DSS). Aquellos primeros data warehouses necesitaban un volumen enorme de redundancias.

En la actualidad, la inteligencia artificial y el aprendizaje automático están transformando casi todos los sectores La expansión de big data y la aplicación de nuevas tecnologías digitales están impulsando un cambio en los requisitos y las capacidades de los data warehouses.

Arquitectura de un data warehouse

La arquitectura de un data warehouse depende de las necesidades específicas de la organización. Algunas de las arquitecturas más comunes son:

- **Sencilla:** todos los data warehouse comparten un diseño básico en el que los metadatos, los datos de resumen y los datos sin procesar se almacenan en el repositorio central del almacén. En uno de los lados, el repositorio se alimenta de fuentes de datos y, en el otro, los usuarios finales acceden para las tareas de análisis, elaboración de informes y extracción.

- **Sencilla con zona de preparación:** es necesario limpiar y procesar los datos operativos antes de colocarlos en el almacén. Aunque esto se puede hacer mediante programación, muchos data warehouse añaden una zona de preparación de datos antes de introducirlos en el almacén, a fin de simplificar la preparación.

- **Radial:** al añadir **data marts** entre el repositorio central y los usuarios finales, las organizaciones pueden personalizar su data warehouse para atender a varias áreas de negocio. Cuando los datos ya están listos para el uso, se mueven al data mart correspondiente.

- **Entorno de pruebas:** los entornos de pruebas (sandbox) son zonas seguras, privadas y protegidas donde las empresas pueden explorar de forma rápida e informal nuevos conjuntos de datos o nuevas maneras de analizarlos sin tener que cumplir con las reglas y los protocolos formales del data warehouse.

Ventajas de un almacén de datos

Los almacenes de datos ofrecen una ventaja exclusiva y global: permitir a las organizaciones **analizar grandes cantidades de datos** de variables **y extraer un valor significativo**, además de mantener un registro histórico.

Los data warehouse cuentan con cuatro características exclusivas:

- **Orientados a sujetos:** pueden analizar datos sobre un tema particular o área funcional (como las ventas).

- **Integrados:** los data warehouse crean uniformidad entre diferentes tipos de datos provenientes de fuentes dispares.

- **No volátiles:** cuando la información entra en un data warehouse, es estable y no cambia.

- **Tienen en cuenta las variaciones a lo largo del tiempo:** la analítica de los data warehouse toma en consideración los cambios que se producen con el tiempo.

Un almacén de datos bien diseñado realizará las consultas muy rápidamente, ofrecerá un alto rendimiento de datos y proporcionará suficiente flexibilidad para que los usuarios finales puedan segmentar y desglosar o reducir el volumen de datos para un examen más detallado con el fin de satisfacer diversas necesidades, tanto generales como específicas.



El data warehouse es la base funcional para los entornos de BI middleware (Inteligencia Empresarial) que suministran informes, cuadros de mando y otras interfaces a los usuarios finales.

Almacenes de datos, data marts y almacenes de datos operativos

Aunque desempeñan funciones similares, los almacenes de datos, los data marts y los almacenes de datos operativos (ODS) tienen diferencias entre sí.

Un **data mart** lleva a cabo las mismas funciones que un almacén de datos, pero con un alcance mucho más limitado (por lo general, para un solo departamento o línea de negocio). Es por eso que es más fácil establecer un data mart que un data warehouse. Sin embargo, suelen introducir incoherencias, ya que es difícil administrar y controlar los datos de manera uniforme en muchos data marts.

Los **ODS** sólo admiten operaciones diarias, por lo que su visión de los datos históricos es muy limitada. Aunque funcionan muy bien como fuentes de datos actuales y los almacenes de datos los emplean de ese modo, no admiten consultas avanzadas de datos históricos.