

DATA WAREHOUSE

1. ¿Qué es un 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 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 vista y a gran velocidad. Se trata de uno de los componentes más importantes de la inteligencia empresarial en el entorno actual en el que operan las empresas.

3. Características del Data Warehouse

Bill Inmon fue uno de los pioneros en tratar el tema relativo al «**Almacén de datos**», traducción literal de este término, y destacó las siguientes **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 él almacenados.
- No volátil. La información no se puede modificar ni eliminar una vez almacenado el dato. Sólo podrá ser usada como lectura o consulta en un futuro.
- Integrado. Los datos almacenados en el **data warehouse** 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.** El **data warehouse** 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

A continuación, detallamos las principales **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 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 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.

6. Data Warehouse en la nube

Los data warehouses están atravesando actualmente dos transformaciones muy importantes que tienen el potencial de impulsar niveles significativos de innovación empresarial:

- **La primera área de transformación es el impulso para aumentar la agilidad general.** La gran mayoría de los departamentos de TI están experimentando un rápido aumento de la demanda de datos. Los directivos quieren tener acceso a más y más datos históricos, mientras que, al mismo tiempo, los científicos de datos y los analistas de negocios están explorando formas de introducir nuevos flujos de datos en el almacén para enriquecer el análisis existente, así como impulsar nuevas áreas de análisis. Esta rápida expansión de los volúmenes y fuentes de datos significa que los equipos de TI necesitan invertir más tiempo y esfuerzo asegurando que el rendimiento de las consultas permanezca constante y necesitan proporcionar cada vez más entornos para equipos individuales para validar el valor comercial de los nuevos conjuntos de datos.
- **La segunda área de transformación gira en torno a la necesidad de mejorar el control de costes.** Existe una creciente necesidad de hacer más con cada vez menos recursos, al mismo tiempo que se garantiza que todos los datos sensibles y estratégicos estén completamente asegurados, a lo largo de todo el ciclo de vida, de la manera más rentable.