

## Data Warehouse.

Los centros de información son espacios dedicados a almacenar y procesar datos digitales, es decir, se trata de un sitio físico donde el equipo informático se agrupa en armarios de almacenamiento para formar un sistema de información.

Debido al constante incremento del volumen de información y datos de las empresas, ha surgido la necesidad de implementar sistemas que cubran esta urgencia de almacenamiento y un rápido acceso a los mismos.



## ¿Qué es un Data Warehouse?

Un Data Warehouse es un tipo de **sistema de gestión de datos** diseñado para habilitar y dar soporte a las tareas de inteligencia empresarial (BI), especialmente las analíticas. Se han diseñado para realizar consultas y tareas de análisis, y suelen contener grandes cantidades de datos históricos. A menudo, la información dentro de éste proviene de una amplia gama de fuentes.

**Centraliza y fusiona** grandes cantidades de datos de múltiples fuentes. Gracias a sus capacidades analíticas, las organizaciones pueden obtener información empresarial valiosa a partir de los datos y mejorar las decisiones.

## Características

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.

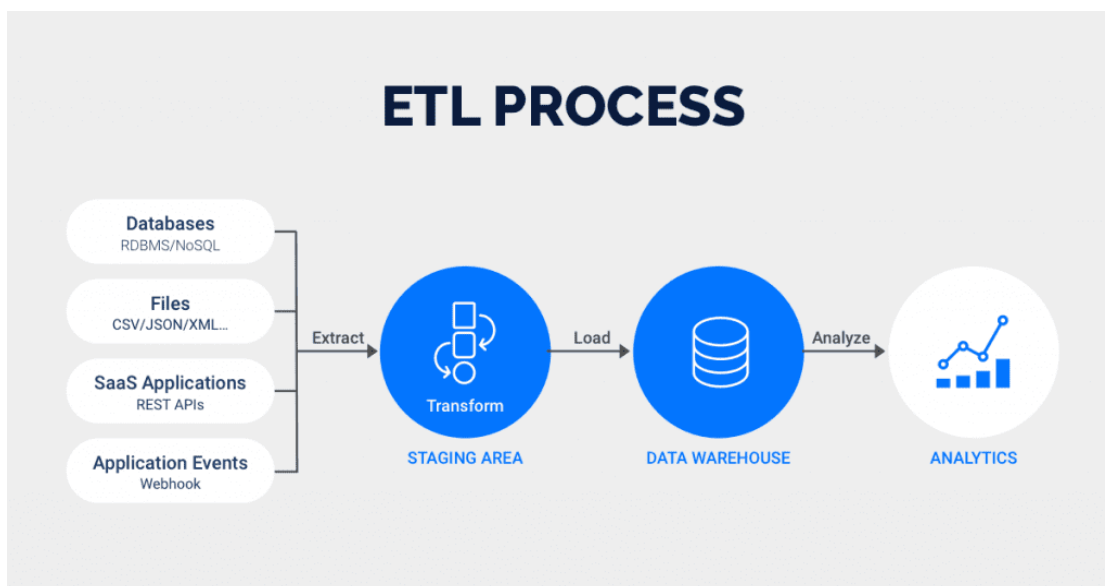
Los data warehouses disponen de las siguientes características:

- **Orientados a sujetos:** Un almacén de datos está orientado a temas, ya que ofrece información sobre un tema en lugar de las operaciones en curso de las empresas. Estos temas pueden ser ventas, marketing, distribuciones, etc.
- **Integrados:** Agrupa en un establecimiento de una unidad de medida común todos los datos similares de distintas bases de datos.
- **No volátiles:** los datos anteriores no se borran cuando se ingresan nuevos datos. Los datos son de solo lectura y se actualizan periódicamente.
- **Variante de tiempo:** Los datos recopilados en un almacén de datos se reconocen con un período particular, contiene un elemento tiempo.

## Procesos de un DW

Los procesos de extracción, transformación y carga (ETL) son importantes ya que son la forma en que los datos se guardan en una base de datos. Implican las siguientes operaciones:

- **Extracción:** obtener la información deseada a partir de los datos almacenados en fuentes externas.
- **Transformación:** operación sobre los datos para que puedan ser cargados en el data warehouse o se puedan migrar de éste a otra base de datos.
- **Carga:** consiste en almacenar los datos en la base de datos final, por ejemplo el almacén de datos objetivo normal.



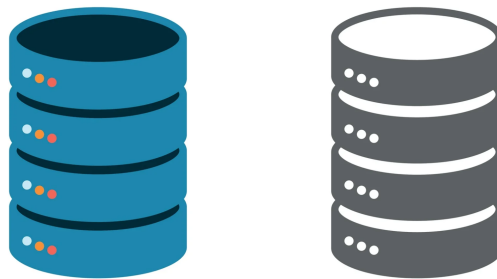
## Beneficios de su uso

Los beneficios de usar un datawarehouse son:

- Facilidad de acceso a la información
- Mayor flexibilidad y rapidez de respuesta
- Consolidación y homogeneización de la información
- Mejor comunicación entre departamentos de la empresa
- Entrega de información relevante que antes no se almacenaba
- Se tiene una base confiable para aplicar técnicas de analytics

## Arquitectura

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 warehouses 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 warehouses 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.