

Data warehouse

Índice

- 01 [Data warehouse](#)
- 02 [Implentación](#)
- 03 [Conclusiones](#)



01

Data warehouse

¿Qué es?

El **data warehouse** un repositorio de datos que:

- Posee muy fácil acceso.
- Consolida muchas fuentes de datos, en un único almacén de datos.
- Transforma los datos en grupos de información sobre temas específicos de negocios, por ejemplo ventas, logística, producción, etc.
- Permite realizar fácilmente consultas, análisis, reportes para la toma de decisiones.

Objetivos

El **data warehouse** busca:

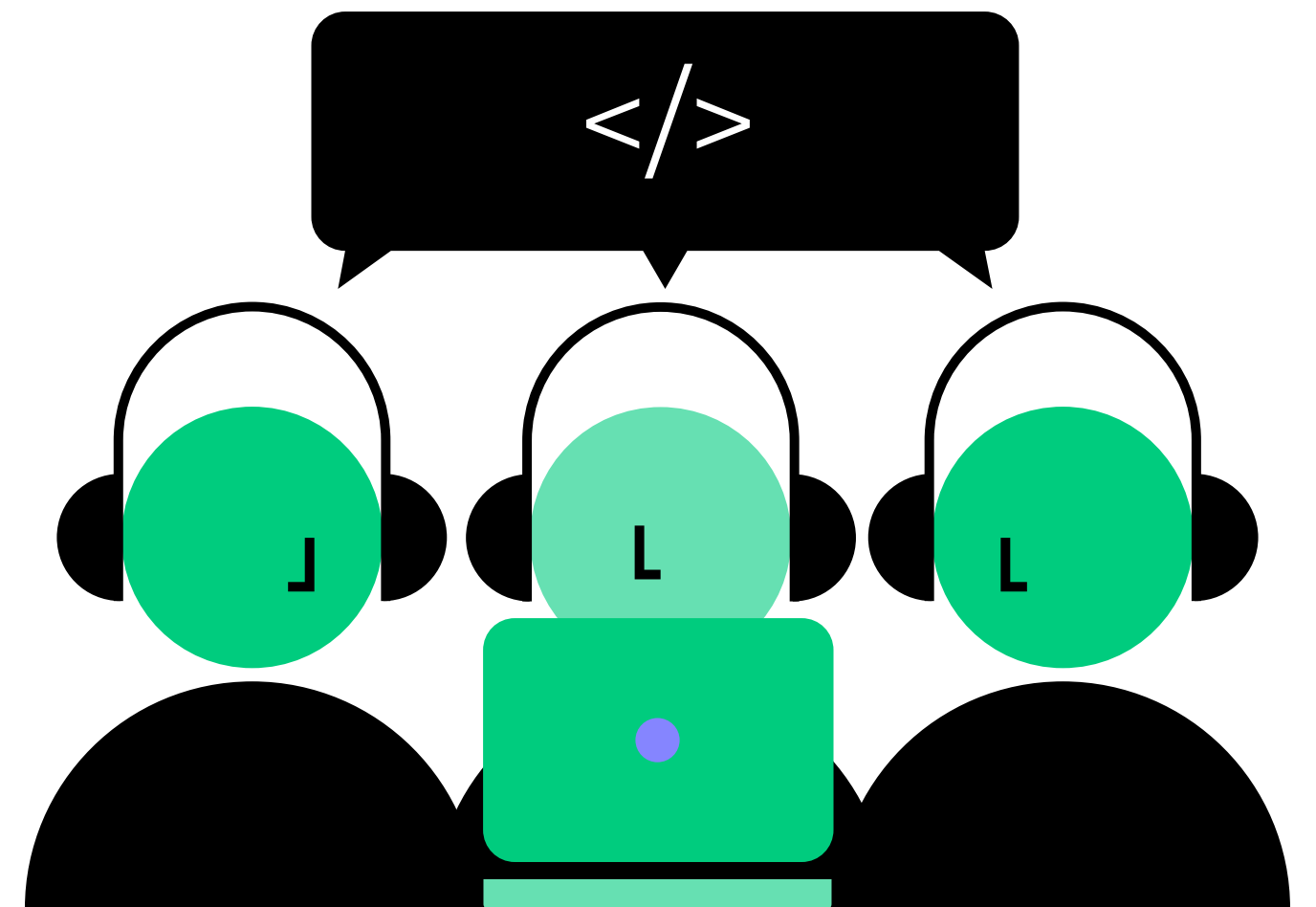
- **Tener información accesible:** los contenidos del data warehouse son entendibles, navegables y de rápido acceso ¡Es muy fácil hacer consultas!
- **Tener información consistente:** es decir, con un único significado para todos.
- **Tener información adaptable y elástica:** el data warehouse está diseñado para cambios continuos.
- **Proteger la información** y el gobierno de datos.
- **Obtener información** para la toma de decisiones.

Un datawarehouse nos tiene que
brindar una visión unificada y
completa de los datos a lo que
comúnmente nos referiremos
como visión 360°.



Características

El **data warehouse** tiene como característica ofrecer los datos unificados en un único lugar que representen a las mismas entidades de una manera unificada, sin importar el origen de esos datos.



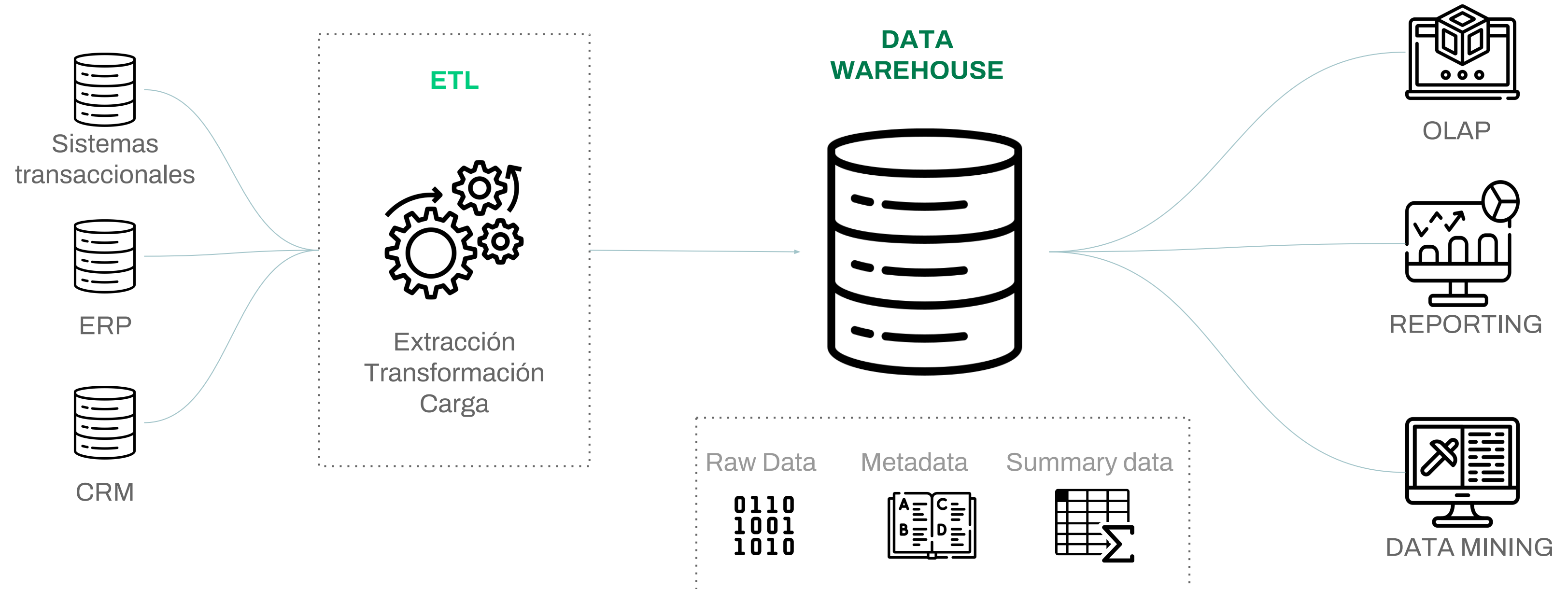
Diferencias con transaccional

Transaccional	Data warehouse
Sistemas fuentes (ERP, CRM, E-commerce).	Data warehouse
Los modelos de datos están orientados a ser eficientes y seguros en la gestión de las transacciones (ventas, compras, altas de clientes, etc.).	El modelo de datos de un data warehouse está orientado y focalizado en ayudarnos a integrar los datos , unificarlos y acceder de manera ágil y simple a través de consultas SQL o desde las herramientas de visualización.
Prioriza la escritura de datos . Garantiza el registro de los datos.	Prioriza la lectura de datos . Agiliza la consulta de los datos.

02

Implementación

Implementación del data warehouse



La arquitectura de un Datawarehouse comprende 4 componentes principales

01

Aplicaciones fuentes de datos.

02

Procesos ETL.

03

La base de datos del data warehouse.

04

Herramientas de visualización y minería de datos.

Fuentes de datos

Las fuentes de datos que alimentan el data warehouse son típicamente sistemas OLTP (o sistemas transaccionales), con los cuáles las empresas realizan el registro de sus transacciones.

Pueden ser:

- **Aplicaciones formales:** ERP, CRM, E-commerce, sistemas legacy desarrollados internamente, sistemas de RRHH, etc.
- **Aplicaciones externas** a la empresa, como por ejemplo redes sociales, Google Analytics, informes de mercado, información de entes de control de una determinada industria (como por ejemplo los entes de control, bancos, compañías de seguros, alimentos, farmacéuticas, impuestos, salud y accidentes de trabajo, etc.).
- **Información que no está formalizada u organizada** en aplicaciones OLTP sino que se almacena en planillas de cálculo archivos CSV, TXT o incluso PDF, entre otros formatos.

Procesos ETL

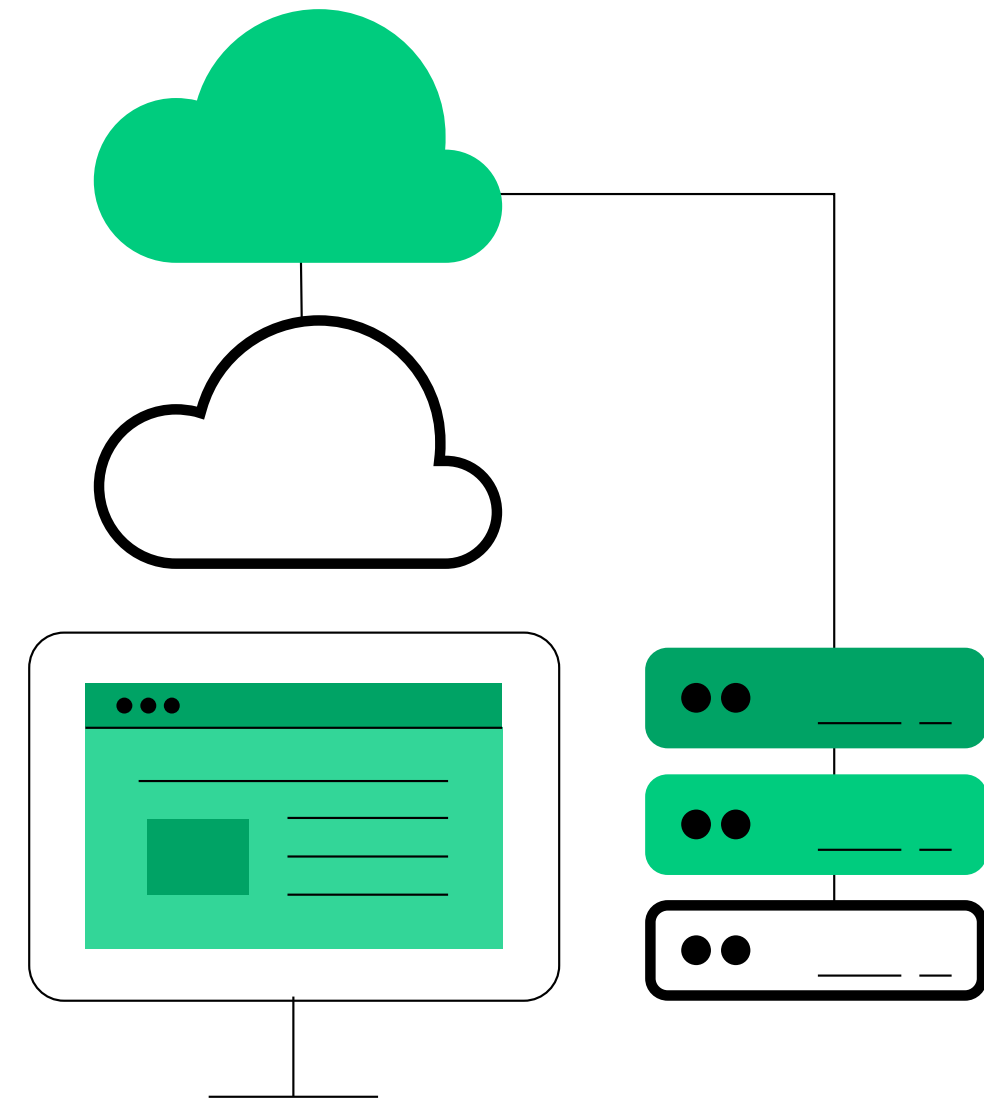
Estos procesos se ejecutan en forma automática y periódica (diariamente, semanalmente, mensualmente, cada hora, etc) y son esenciales para poder mantener la información del data warehouse actualizada, consistente con los sistemas fuentes, unificada y de calidad.

Los procesos ETL (por sus siglas en inglés, E=Extract, T=Transform, L=Load) consisten en procesos Batch (proceso en lote) que:

- Extraen información desde las distintas fuentes de datos.
- Realizan transformaciones simples o complejas (que incluyen normalmente algoritmos para mejorar la calidad de los datos)
- Cargan dichos datos en el data warehouse.

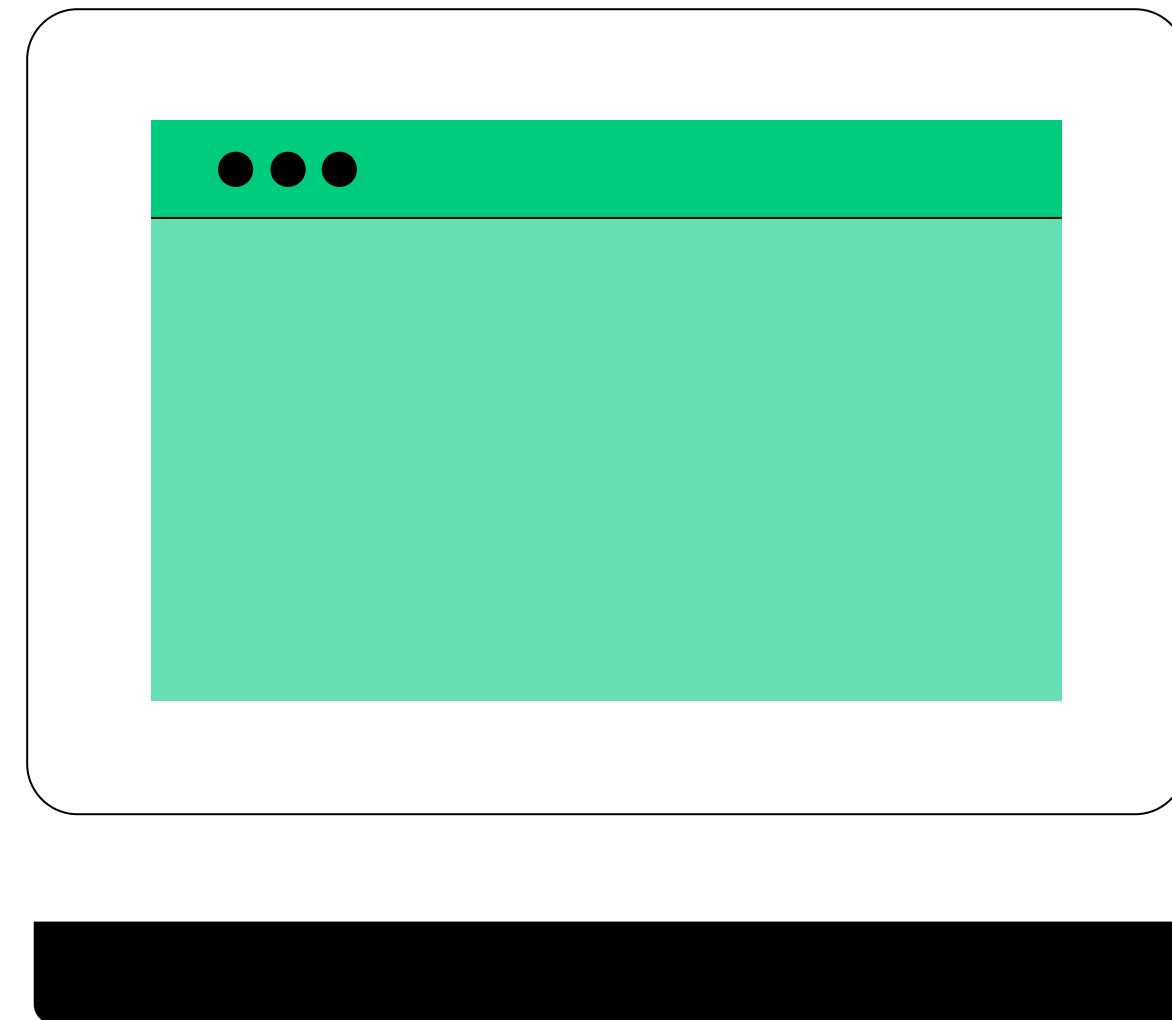
Bases de datos

El tercer componente es la base de datos del data warehouse, un repositorio de datos modelado específicamente para realizar consultas a grandes volúmenes de datos. Es básicamente una base de datos relacional que puede tener un motor optimizado a tal fin, si el volumen de datos así lo requiere.



Herramientas de visualización

El cuarto componente son las herramientas de visualización: estas nos permiten presentar la información. Pueden ser herramientas solo de visualización o que también definan modelos de data mining o de data science más complejos.

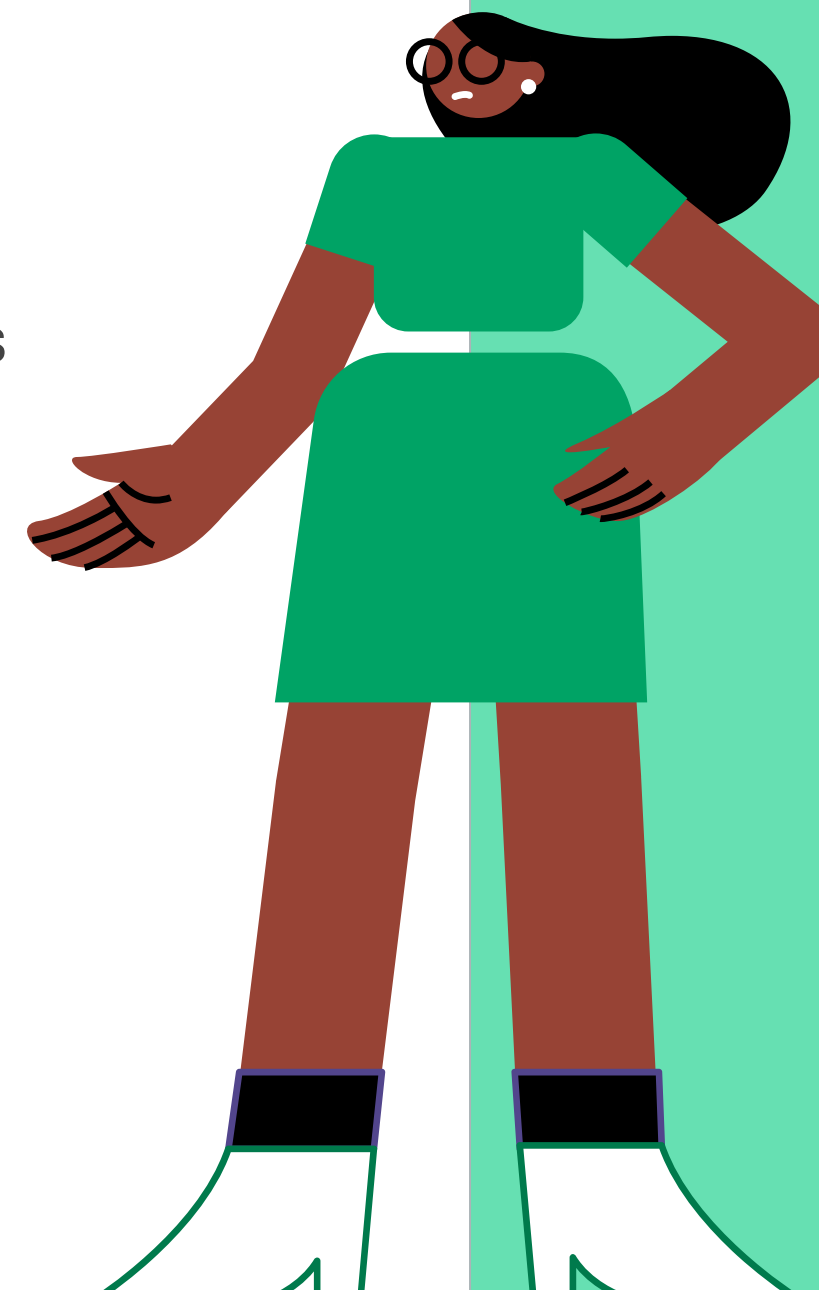


03 Conclusiones

Conclusiones

En resumen, un data warehouse es un repositorio de datos:

- Construido y optimizado para analizar y consultar grandes volúmenes de datos fácilmente.
- Que nos permite tener una visión unificada y con calidad de los datos sin importar el origen del mismo.
- Debe brindarnos una visión completa o de 360° de una empresa, dado que incluye todos los datos necesarios para su gestión.
- Es un componente necesario (pero no suficiente) para asegurar el gobierno de datos, nos permite consolidar una cultura orientada a datos en la organización.



¡Muchas gracias!