

Data lake

Índice

- 01 [Concepto data lake](#)
- 02 [Data lake vs. data warehouse](#)
- 03 [Arquitectura](#)
- 04 [Conclusiones](#)



01

Concepto de data lake



El data lake va más allá del data warehouse, superando algunas de sus limitaciones. Data lake es la siguiente generación en materia de almacenamiento de datos.



Data lake se trata de una forma distinta de orientar la utilización de los datos organizacionales y el reporting. Es la única manera de poder responder a consultas sobre el negocio, haciéndolo sin barreras.

La llegada del data lake no desplaza el **data warehouse**, sino que sirve para ampliar las capacidades de la organización en lo que concierne a su gestión de datos.



Beneficios

01

Facilidad para innovar en la utilización de los datos de la organización, tanto para reporting como para analytics.

02

Mayor agilidad, porque se adapta a la necesidad del usuario respondiendo en el tiempo adecuado, incluso en tiempo real.

03

Más capacidad y flexibilidad en un entorno en el que confluyen datos procedentes de la organización con información proveniente del exterior.

02

Data lake vs. data warehouse

Data lake

Data lake es mejor para gestionar grandes volúmenes de información no estructurada o semi estructurada.

Data warehouse

Data warehouse permite una visión unificada del negocio.

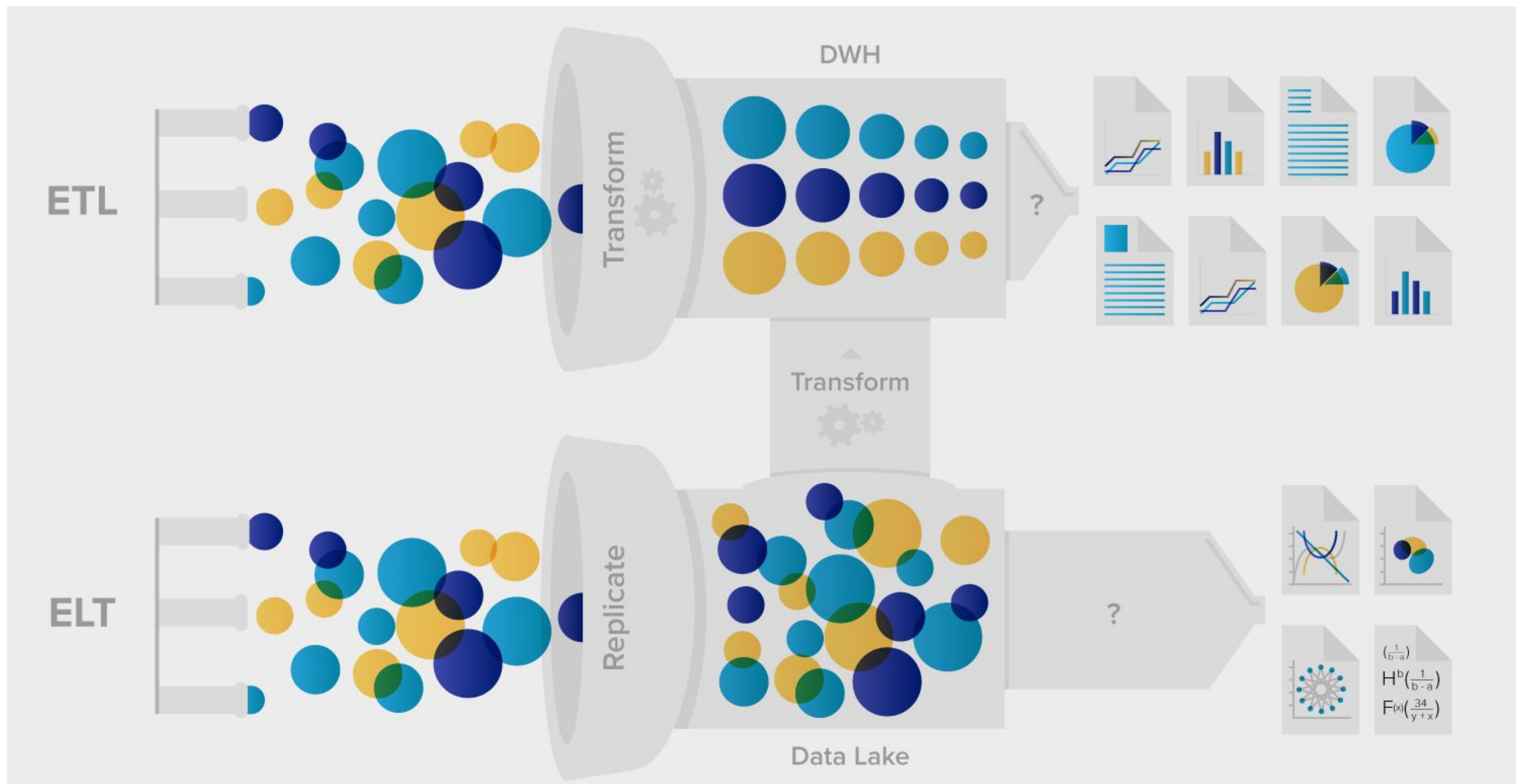


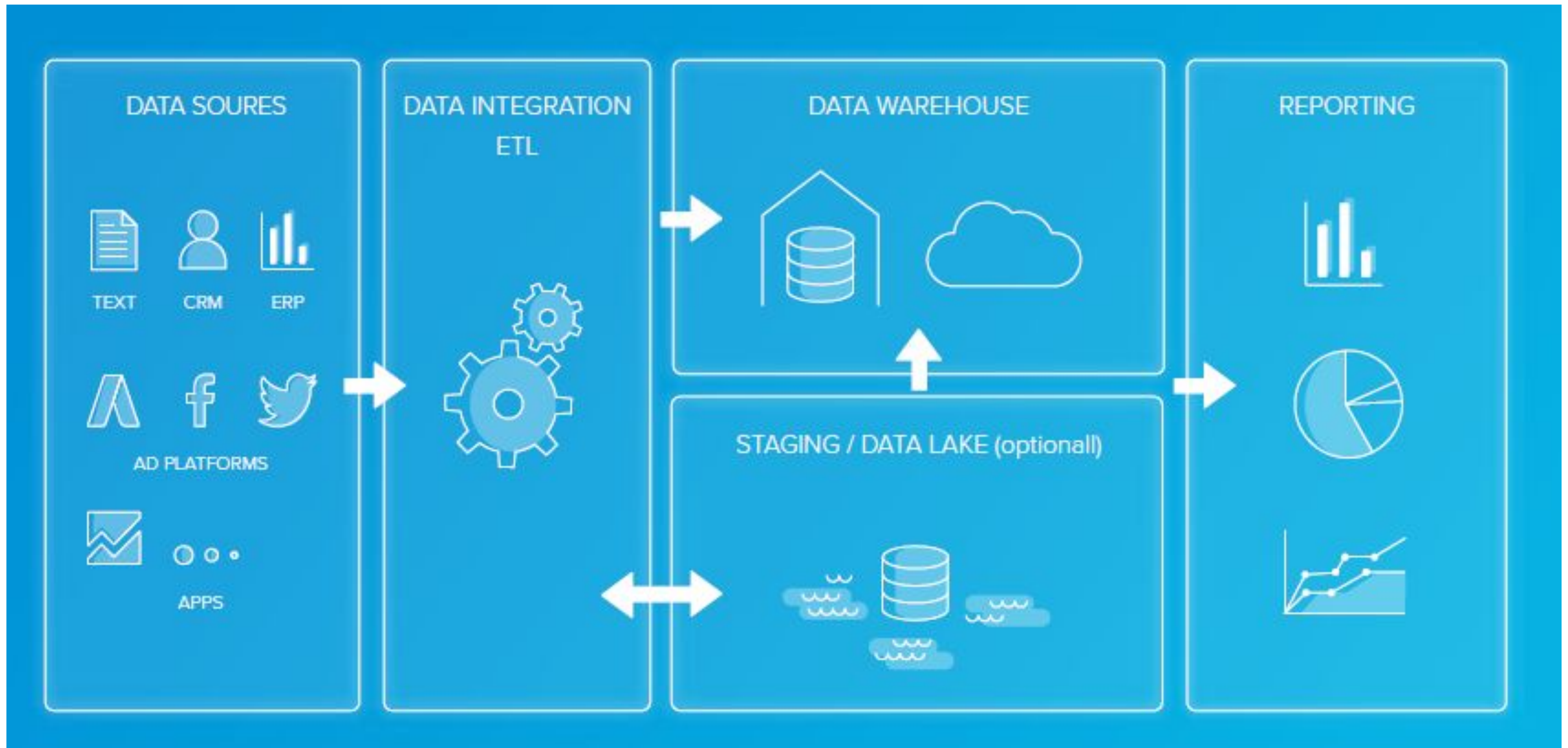
Es posible utilizar ambos. El data lake permitirá a los usuarios **experimentar con diferentes modelos de datos y transformaciones**, antes de configurar un nuevo esquema en un almacén de datos y servirá como área de ensayo, desde la cual suministrar datos al data warehouse. Y, a su vez, este almacén de datos limpios con valor conocido **podrá nutrir y retroalimentar al data lake corporativo**.

Obtener lo mejor de ambas soluciones es cuestión de visión y arquitectura. Y es la manera de no dejar escapar ningún dato valioso.



03 Arquitectura





04

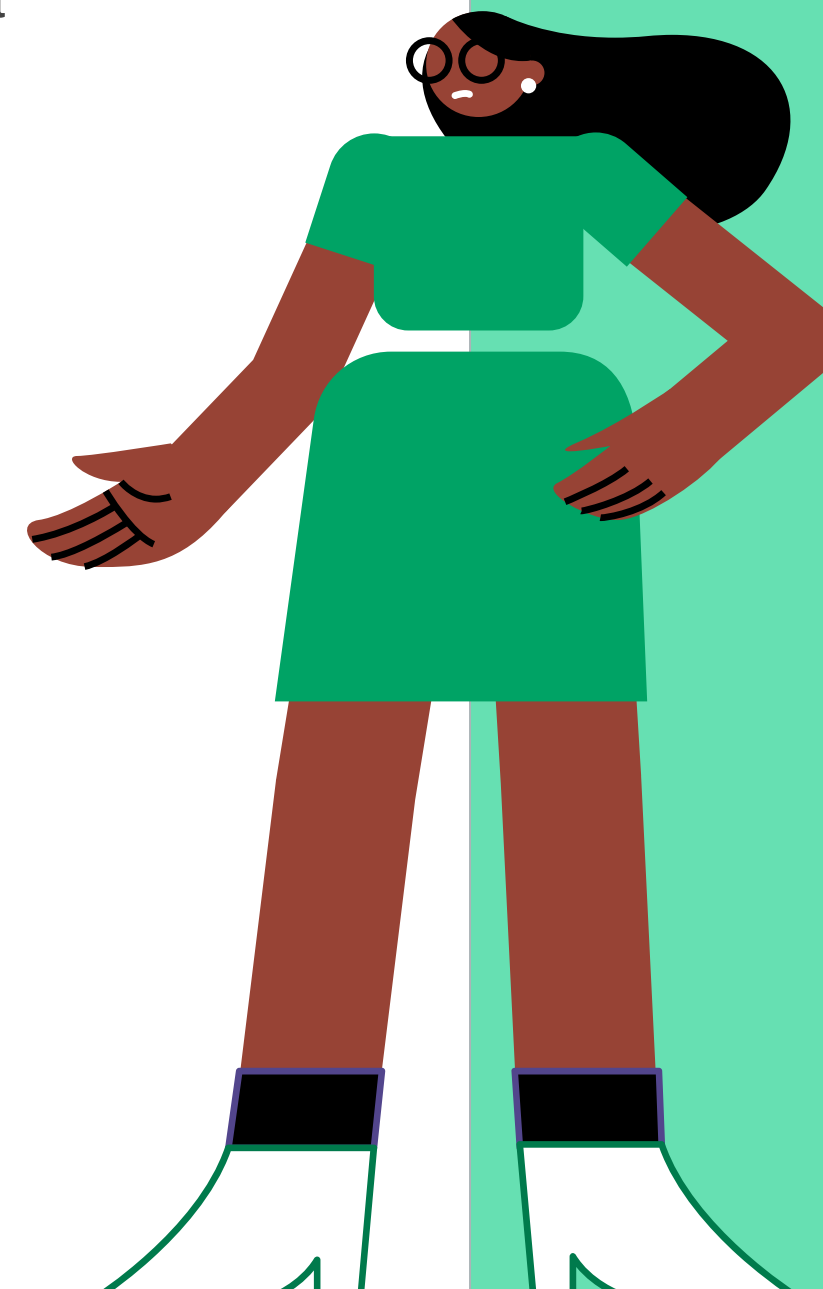
Conclusiones

Conclusiones

El **Data Lake** es el concepto más moderno para centralizar datos para la toma de decisiones. Los esquemas de ingesta de datos que se utilizan en los data lake, pensados para ser utilizados en la nube, resultan muy eficientes para procesar grandes volúmenes de datos. La posibilidad de almacenar datos estructurados, semi estructurados y no estructurados en una misma plataforma, con muy bajo costo es una gran ventaja de los data lakes.

Sin embargo, el **data warehouse** sigue siendo un componente clave para almacenar los datos estructurados más relevantes para la toma de decisiones, por su facilidad para disponer un acceso rápido y seguro.

En una arquitectura moderna, encontraremos al data lake y al data warehouse integrados, y combinando sus ventajas.



¡Muchas gracias!