Innovations in digital technology are generating more data from a greater variety of sources each day. You don't want to lose it, but where do you keep it while exploring the potential value? Data lakes and warehouses are complementary solutions to your data storage and analytics needs.
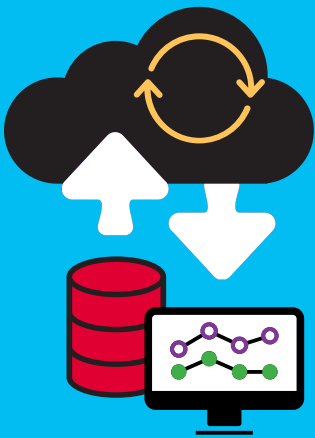
DATA LAKE VS. DATA WAREHOUSE

Extract, load, transform

Data lakes follow an ELT process. Data is gathered from multiple sources, loaded, and stored in its raw format. It's then transformed from within on the fly as needed.

Extract, transform, load

Data warehouses work back to front. The data structure is predetermined first. An ETL tool then cleans and transforms data to fit before it passes into the warehouse.

Store all data types.

Not limited to structured data, data lakes can store semi-structured logs or clickstreams and unstructured data like images, video, and audio.

Store structured data.

Data flows into a warehouse from transactional systems and business applications. Think ERP and CRM. It consolidates structured data into a single source of truth.
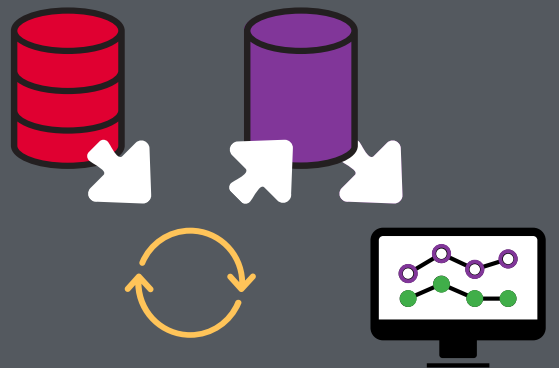
Ingest quickly.

Data collection is justified hoarding, and data lakes support the habit. They allow businesses to pull from more sources in less time yet catalog and secure the assests.

Query quickly.

Data warehouses prove the time-saving value of organization. Data is sorted into schemas and the most frequently used data sits in the fastest storage devices for rapid querying.

101010101010101

As your data grows, a combination may be best. You might gather and prepare assests in the data lake, then move them into the warehouse.

1010101010

Focus on the future.

Data lakes are typically used by business analysts and data scientists who leverage machine learning to profile and predict trends.

Focus on the now.

Data warehouses are specially designed for business analysts to perform operational reporting and analysis on current and historic data.