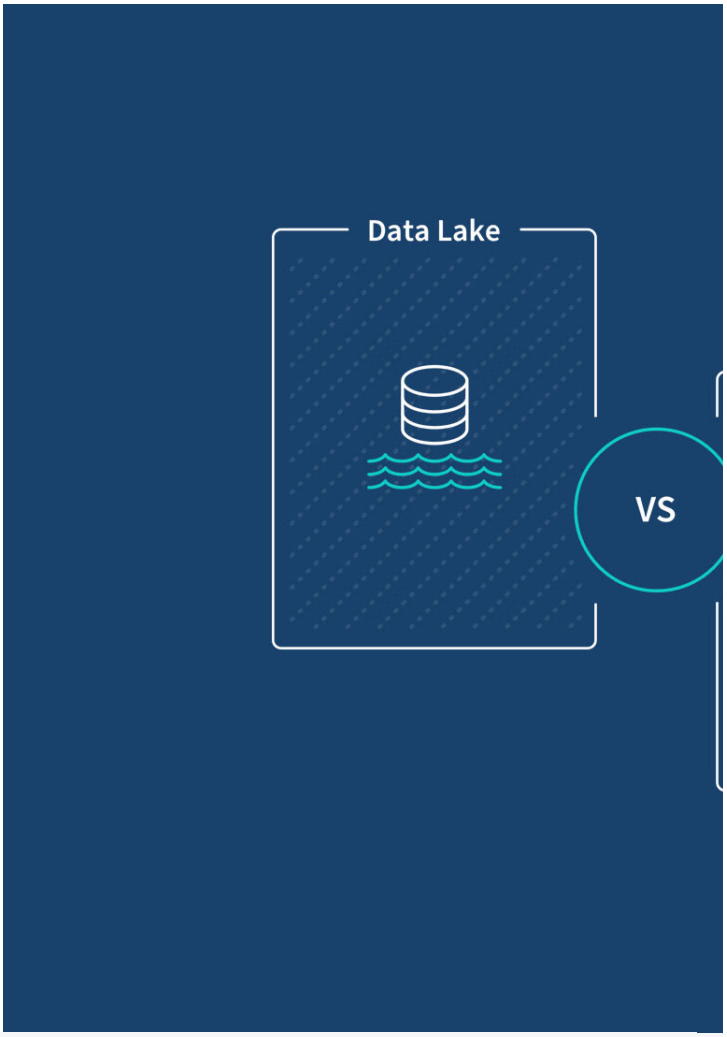
Data Lake vs Data Warehouse

What are the key benefits and differences? This guide provides definitions and practical advice to help you understand the differences as you evaluate data lake vs data warehouse for your organization.

Get Comparison Guide



DATA LAKE VS DATA WAREHOUSE GUIDE

A data lake is a massive repository of structured and unstructured data, and the purpose for this data has not been defined. A data warehouse is a repository of highly structured historical data which has been processed for a defined purpose.

What is a Data Lake?

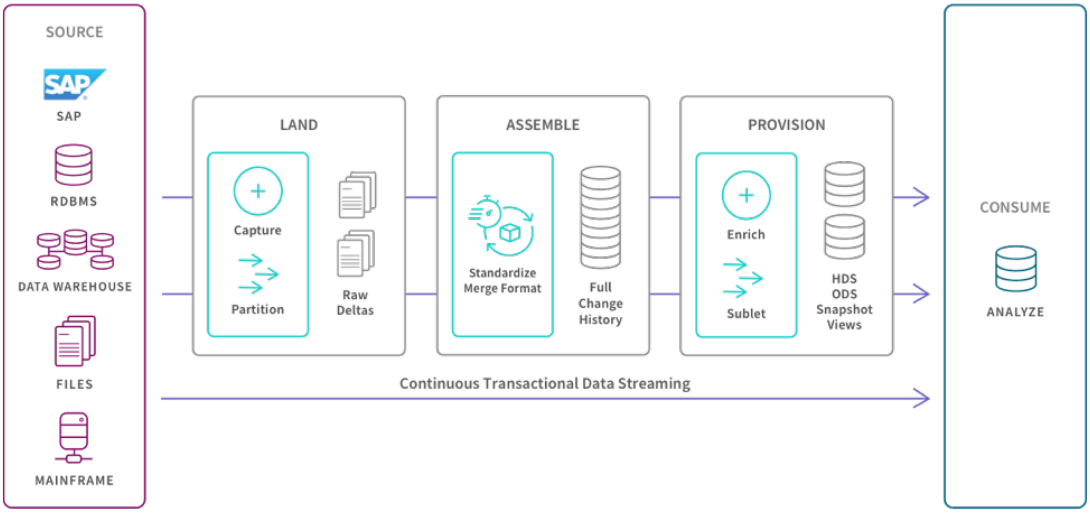
A data lake is a repository that stores all of your organization's data — both structured and unstructured. Think of it as a massive storage pool for data in

its natural, raw state (like a lake). A data lake architecture can handle the huge volumes of data that most organizations produce without the need to structure it first. Data stored in a data lake can be used to build data pipelines to make it available for data analytics tools to find insights that inform key business decisions.

Data Lake Benefits

Because the large volumes of data in a data lake are not structured before being stored, skilled data scientists or end-to-end self-service-BI tools can gain access to a broader range of data far faster than in a data warehouse.

1. Massive volumes of structured and unstructured data like ERP transactions and call logs can be stored cost effectively.
2. Data is available for use far faster by keeping it in a raw state.
3. A broader range of data can be analyzed in new ways to gain unexpected and previously unavailable insights.



Your data engineers can build ETL data pipelines and schema-on-read transformations to make data stored in a data lake available for analytics, data science, and machine learning. Managed data lake creation tools help you overcome the limitations of slow, hand-coded scripts and scarce engineering resources.

Today, many companies are adopting **Delta Lake**, an open-source storage layer that leverages ACID compliance from transactional databases to

layer that leverages ACID compliance from transactional databases to enhance reliability, performance, and flexibility in data lakes. It's particularly useful for scenarios requiring transactional capabilities and schema enforcement within your data lake. It enables the creation of data lakehouses, which support both data warehousing and machine learning directly on the data lake. It offers features such as scalable metadata handling, data versioning, and schema enforcement for large-scale datasets, ensuring data quality and reliability for analytics and data science tasks.

Learn More About Data Lakes ›

Data Lake or Data Warehouse?

Back

Technology Partners and Integrations

Extend the value of Qlik data integration and analytics

Back

Back

All Products

Back

Back

Solution Partners

Back

Back

Data Integration and Quality Pricing

Rapidly deliver trusted data to drive smarter decisions with the right data integration plan.

Analytics Pricing

Deliver better insights and outcomes with the right analytics plan.

AI/ML Pricing

Build and deploy predictive AI apps with a no-code experience.

Back

Customer Stories

Back

Events & Webinars

Back

Resource Library

Back

Glossary

Back

	Data Lake	Data Warehouse
1. Data Storage	A data lake contains all an organization's data in a raw, unstructured form, and can store the data indefinitely — for immediate or future use.	A data warehouse contains structured data that has been cleaned and processed, ready for strategic analysis based on predefined business needs.
2. Users	Data from a data lake — with its large volume of unstructured data — is typically used by data scientists and engineers who prefer to study data in its raw form to gain new, unique business insights.	Data from a data warehouse is typically accessed by managers and business-end users looking to gain insights from business KPIs, as the data has already been structured to provide answers to pre-determined questions for analysis.
3. Analysis	Predictive analytics, machine learning, data	Data visualization, BI, data analytics.

	visualization, BI, big data analytics.	
4. Schema	Schema is defined after the data is stored in a data lake vs data warehouse, making the process of capturing and storing the data faster.	In a data warehouse, the schema is defined before the data is stored. This lengthens the time it takes to process the data, but once complete, the data is at the ready for consistent, confident use across the organization.
5. Processing	ELT (Extract, Load, Transform). In this process, the data is extracted from its source for storage in the data lake, and structured only when needed.	ETL (Extract, Transform, Load). In this process, data is extracted from its source(s), scrubbed, then structured so it's ready for business-end analysis.
6. Cost	Storage costs are fairly inexpensive in a data lake vs data warehouse. Data lakes are also less time-consuming to manage, which reduces operational costs.	Data warehouses cost more than data lakes, and also require more time to manage, resulting in additional operational costs.

Data Lake ROI: 5 Principles for Managing Data Lake Pipeline

Download Whitepaper

DataOps for Analytics

Modern data integration delivers real-time, analytics-ready and actionable data to any analytics environment, from Qlik to Tableau, Power BI and beyond.

Real-time data streaming (CDC)

Extend enterprise data into live streams to enable modern analytics and microservices with a simple, real-time, and comprehensive solution.

Explore Data Streaming ›

Agile data warehouse automation

Quickly design, build, deploy and manage purpose-built cloud data warehouses without manual coding.

Explore Data Warehouse Automation ›

Managed data lake creation

Automate complex ingestion and transformation processes to provide continuously updated and analytics-ready data lakes.

Explore Data Lake Creation ›

Learn more about data integration with Qlik

Try for Free

Contact Us

Why Qlik?

- Why Qlik
- Trust and Security
- Trust and Privacy
- Trust and AI
- Why Qlik for AI
- Compare Qlik
- Featured Technology Partners
- Data Sources and Targets

About Qlik

- Company
- Leadership
- CSR
- DEI&B
- Academic Program
- Partner Program
- Careers
- Newsroom
- Global Office/Contact

Products

DATA INTEGRATION AND QUALITY

- Qlik Talend
- Qlik Talend Cloud
- Talend Data Fabric

ANALYTICS & AI

- Qlik Cloud Analytics
- Qlik Answers
- Qlik Predict
- Qlik Automate

Pricing

- Data Integration Pricing
- Analytics Pricing
- AI/ML Pricing

Solutions

INDUSTRIES

- ISV
- Financial Services
- Healthcare
- Public Sector/Government
- US Government
- Retail
- Communications
- Manufacturing
- Consumer Products
- Energy Utilities
- High Tech
- Life Sciences

BY ROLE

- Sales
- Marketing
- Finance
- Operations
- Product Intelligence
- HR & People
- IT

SOLUTION PARTNERS

- Find a Partner
- Global SIs

Learn

- Blog
- Customer Stories
- Events
- Glossary
- Community
- Training

RESOURCE CENTER

- Resource Library
- Analysts Reports
- Whitepapers & Ebooks
- Webinars
- Videos
- Datasheet & Brochures
- Customer Stories

Support

- Community
- Support
- Customer Portal
- Onboarding
- Product Documentation
- Training

Qlik

Qlik Community