What is a Data Lake?

A data lake is a centralized repository designed to store, process, and secure large amounts of structured, semistructured, and unstructured data. It can store data in its native format and process any variety of it, ignoring size limits.

Learn more about modernizing your data lake (https://cloud.google.com/solutions/data-lake) on Google Cloud.

Ready to get started? New customers get $300 in free credits (https://console.cloud.google.com/freetrial) to spend on Google Cloud.
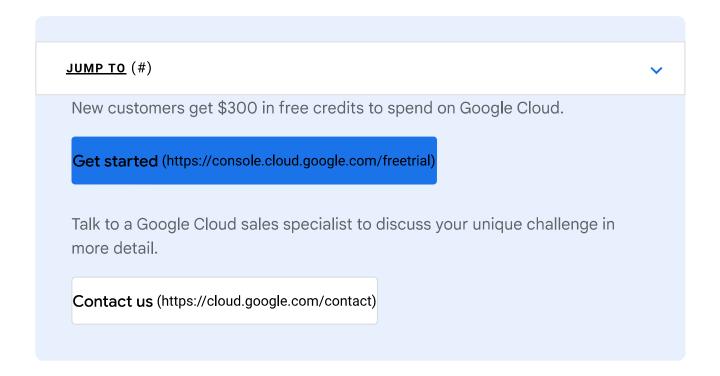
Data lake overview

A data lake provides a scalable and secure platform that allows enterprises to: ingest any data from any system at any speed—even if the data comes from on-premises, cloud, or edge-computing systems; store any type or volume of data in full fidelity; process data in real time or batch mode; and analyze data using SQL, Python, R, or any other language, third-party data, or analytics application.

Data lake vs. data warehouse: A data lake is also defined by what it isn't. It's not just storage, and it's not the same as a data warehouse.

While data lakes and data warehouses all store data in some capacity, each is optimized for different uses. Consider them complementary rather than competing tools, and

companies might need both. As a point of comparison, data warehouses are often ideal for the kind of repeatable reporting and analysis that's common in business practices, such as monthly sales reports, tracking of sales per region, or website traffic.

JUMP TO (#) ⌄

New customers get $300 in free credits to spend on Google Cloud.

Get started (https://console.cloud.google.com/freetrial)

Talk to a Google Cloud sales specialist to discuss your unique challenge in more detail.

Contact us (https://cloud.google.com/contact)

Do you need a data lake?

When determining if your company needs a data lake, keep in mind the types of data you're working with, what you want to do with the data, the complexity of your data acquisition process, and your strategy for data management and governance, as well as the tools and skill sets that exist in your organization.

Companies today are also starting to look at the value of data lakes through a different lens—a data lake isn't only about storing full-fidelity data. It's also about users gaining a deeper understanding of business situations because they have more context than ever before, allowing them to accelerate analytics experiments.

Developed primarily to handle large volumes of big data, companies can typically move raw data via batch and/or stream into a data lake without transforming it. Enterprises rely on data lakes in key ways to help:

- Lower the total cost of ownership

- Simplify data management

- Prepare to incorporate artificial intelligence and machine learning

- Speed up analytics

- Improve security and governance

Because data lakes provide the foundation for analytics and artificial intelligence, businesses across every industry are using them to increase revenue, save money, and reduce risk.

Media and entertainment

A company offering streaming music, radio, and podcasts can increase revenue by improving their recommendation system, so users consume their service more, allowing the company to sell more ads.

Telecommunications

A multinational telecommunications company can save money by building churn-propensity models that reduce customer churn.

Financial services

An investment firm can rely on data lakes to power machine learning, so they can manage portfolio risks as soon as real-time market data is made accessible.
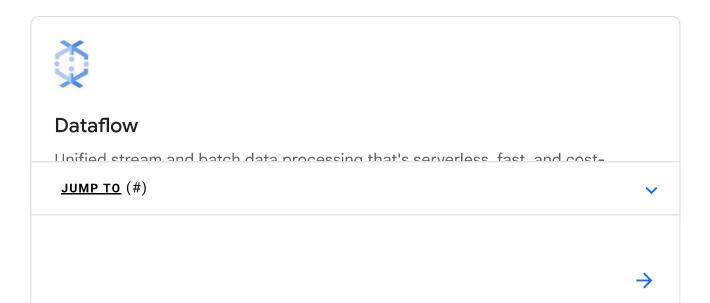
Related products and services

Google Cloud offers a suite of autoscaling services that enable you to build a data lake that integrates with your existing applications, skills, and IT investments. This includes Dataflow (https://cloud.google.com/dataflow/) and Cloud Data Fusion (https://cloud.google.com/data-fusion) for data ingestion, Cloud Storage (https://cloud.google.com/storage) for storage, and Dataproc (https://cloud.google.com/dataproc/) and BigQuery (https://cloud.google.com/bigquery/) for data and analytics processing.

Dataflow

Unified stream and batch data processing that's serverless, fast, and cost-

→

BigQuery

Serverless, highly scalable, and cost-effective cloud data warehouse designed for business agility.

→

Cloud Data Fusion

A fully managed, cloud-native data integration service that helps users efficiently build and manage ETL/ELT data pipelines.

→

Dataproc

Dataproc makes open source data and analytics processing fast, easy, and more

→

Cloud Storage

Globally unified, scalable, and highly durable object storage for developers and enterprises.

→

SOLUTION

Data Lake Modernization

Google Cloud's data lake allows you to securely and cost-effectively ingest, store, and analyze large volumes of diverse, full-fidelity data.

→

SOLUTION

Smart Analytics

Google Cloud's fully managed serverless analytics platform empowers your business while eliminating constraints of scale, performance, and cost.

→

⌄

Take the next step

Start building on Google Cloud with $300 in free credits and 20+ always free products.

Get started for free (https://console.cloud.google.com/freetrial/)

Need help getting started?

Contact sales (https://cloud.google.com/contact/)

Work with a trusted partner

Find a partner (https://cloud.google.com/find-a-partner/)

Continue browsing

See all products (https://cloud.google.com/products/)