

[Get Started for Free](#)[Contact Us](#)

## Analytics on AWS ▾

[Overview](#)[Modern Data Architecture](#)[What is a data lake?](#)[Data Lakes](#)[Partners](#)[Customers](#)[Migrations](#)

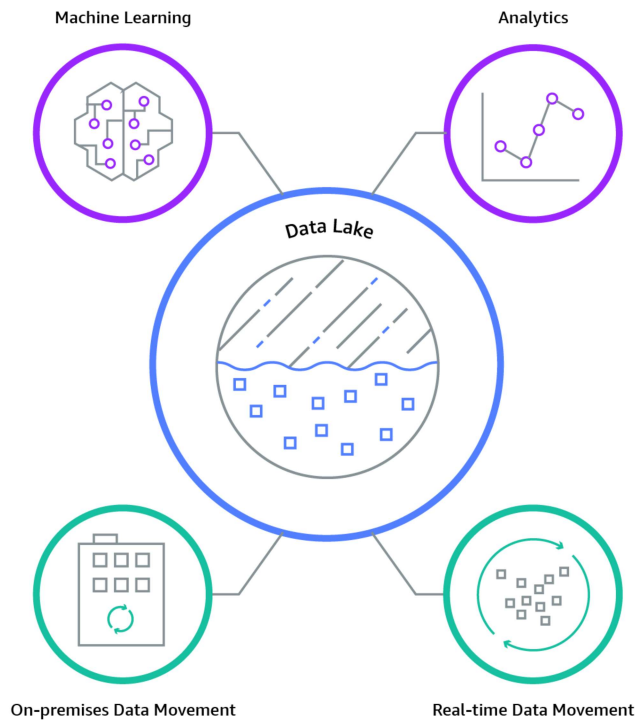
# What is a data lake?

Store all your data in one centralized repository at any scale

[Learn about data lakes and analytics on AWS](#)

## What is a data lake?

A data lake is a centralized repository that allows you to store all your structured and unstructured data at any scale. You can store your data as-is, without having to first structure the data, and run different types of analytics—from dashboards and visualizations to big data processing, real-time analytics, and machine learning to guide better decisions.



## Why do you need a data lake?

Organizations that successfully generate business value from their data, will outperform their peers. An [Aberdeen survey](#) saw organizations who implemented a Data Lake outperforming similar companies by 9% in organic revenue growth. These leaders were able to do new types of analytics like machine learning over new sources like log files, data from click-streams, social media, and internet connected devices stored in the data lake. This helped them to identify, and act upon opportunities for business growth faster by attracting and retaining customers, boosting productivity, proactively maintaining devices, and making informed decisions.

## Data Lakes compared to Data Warehouses – two different approaches

Depending on the requirements, a typical organization will require both a data warehouse and a data lake as they serve different needs, and use cases.

A data warehouse is a database optimized to analyze relational data coming from transactional systems and line of business applications. The data structure, and schema are defined in advance to optimize for fast SQL queries, where the results are typically used for operational reporting and analysis. Data is cleaned, enriched, and transformed so it can act as the “single source of truth” that users can trust.

A data lake is different, because it stores relational data from line of business applications, and non-relational data from mobile apps, IoT devices, and social media. The structure of the data or schema is not defined when data is captured. This means you can store all of your data without careful design or the need to know what questions you might need answers for in the future. Different types of analytics on your data like SQL queries, big data analytics, full text search, real-time analytics, and machine learning can be used to uncover insights.

As organizations with data warehouses see the benefits of data lakes, they are evolving their warehouse to include data lakes, and enable diverse query capabilities, data science use-cases, and advanced capabilities for discovering new information models. Gartner names this evolution the “Data Management Solution for Analytics” or “[DMSA](#).”

Characteristics	Data Warehouse	Data Lake
<b>Data</b>	Relational from transactional systems, operational databases, and line of business applications	Non-relational and relational from IoT devices, web sites, mobile apps, social media, and corporate applications
<b>Schema</b>	Designed prior to the DW implementation (schema-on-write)	Written at the time of analysis (schema-on-read)
<b>Price/Performance</b>	Fastest query results using higher cost storage	Query results getting faster using low-cost storage
<b>Data Quality</b>	Highly curated data that serves as the central version of the truth	Any data that may or may not be curated (ie. raw data)

<b>Users</b>	Business analysts	Data scientists, Data developers, and Business analysts (using curated data)
<b>Analytics</b>	Batch reporting, BI and visualizations	Machine Learning, Predictive analytics, data discovery and profiling

## The essential elements of a Data Lake and Analytics solution

As organizations are building Data Lakes and an Analytics platform, they need to consider a number of key capabilities including:

### Data movement

Data Lakes allow you to import any amount of data that can come in real-time. Data is collected from multiple sources, and moved into the data lake in its original format. This process allows you to scale to data of any size, while saving time of defining data structures, schema, and transformations.

### Securely store, and catalog data

Data Lakes allow you to store relational data like operational databases and data from line of business applications, and non-relational data like mobile apps, IoT devices, and social media. They also give you the ability to understand what data is in the lake through crawling, cataloging, and indexing of data. Finally, data must be secured to ensure your data assets are protected.

### Analytics

Data Lakes allow various roles in your organization like data scientists, data developers, and business analysts to access data with their choice of analytic tools and frameworks. This includes open source frameworks such as Apache Hadoop, Presto, and Apache Spark, and commercial offerings from data warehouse and business intelligence vendors. Data Lakes allow you to run analytics without the need to move your data to a separate analytics system.

### Machine Learning

Data Lakes will allow organizations to generate different types of insights including reporting on historical data, and doing machine learning where models are built to forecast likely outcomes, and suggest a range of prescribed actions to achieve the optimal result.

## The value of a Data Lake

The ability to harness more data, from more sources, in less time, and empowering users to collaborate and analyze data in different ways leads to better, faster decision making. Examples where Data Lakes have added value include:

### Improved customer interactions

A Data Lake can combine customer data from a CRM platform with social media analytics, a marketing platform that includes buying history, and incident tickets to empower the business to understand the most profitable customer cohort, the cause of customer churn, and the promotions or rewards that will increase loyalty.

### Improve R&D innovation choices

A data lake can help your R&D teams test their hypothesis, refine assumptions, and assess results—such as choosing the right materials in your product design resulting in faster performance, doing genomic research leading to more effective medication, or understanding the willingness of customers to pay for different attributes.

### Increase operational efficiencies

The Internet of Things (IoT) introduces more ways to collect data on processes like manufacturing, with real-time data coming from internet connected devices. A data lake makes it easy to store, and run analytics on machine-generated IoT data to discover ways to reduce operational costs, and increase quality.

## The challenges of Data Lakes

The main challenge with a data lake architecture is that raw data is stored with no oversight of the contents. For a data lake to make data usable, it needs to have defined mechanisms to catalog, and secure data. Without these elements, data cannot be found, or trusted resulting in a "data swamp." Meeting the needs of wider audiences require data lakes to have governance, semantic consistency, and access controls.

## Deploying Data Lakes in the cloud

Data Lakes are an ideal workload to be deployed in the cloud, because the cloud provides performance, scalability, reliability, availability, a diverse set of analytic engines, and massive economies of scale. [ESG research](#) found 39% of respondents considering cloud as their primary deployment for analytics, 41% for data warehouses, and 43% for Spark. The top reasons customers perceived the cloud as an advantage for Data Lakes are better security, faster time to deployment, better availability, more frequent feature/functionality updates, more elasticity, more geographic coverage, and costs linked to actual utilization.

## Build your Data Lakes in the cloud on AWS

AWS provides the most secure, scalable, comprehensive, and cost-effective portfolio of services that enable customers to build their data lake in the cloud, analyze all their data, including data from IoT devices with a variety of analytical approaches including machine learning. As a result, there are more organizations running their data lakes and analytics on AWS than anywhere else with customers like NETFLIX, Zillow, NASDAQ, Yelp, iRobot, and FINRA trusting AWS to run their business critical analytics workloads. [Learn more.](#)

## More resources on the data lake

Learn more about data lakes from industry analysts.

Aberdeen: Angling for Insight in Today's Data Lake



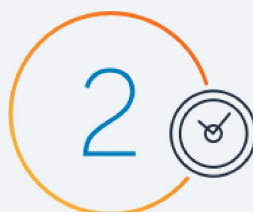
ESG: Embracing a Data-centric Culture Anchored by a Cloud Data451: The Cloud-Based Approach From Big Data



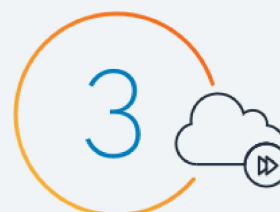
## Get started with AWS



Sign up for an AWS account



Build a secure data lake in days



Start building with AWS

## Get started with data lakes on AWS

[Learn about Data Lakes and Analytics on AWS](#)

[Deploy a data lake with AWS Quick Starts](#)

Have more questions?

[Contact us](#)

[Sign In to the Console](#)

### Learn About AWS

[What Is AWS?](#)

[What Is Cloud Computing?](#)

[AWS Diversity, Equity & Inclusion](#)

[What Is DevOps?](#)

[What Is a Container?](#)

[What Is a Data Lake?](#)

[AWS Cloud Security](#)

[What's New](#)

[Blogs](#)

[Press Releases](#)

### Resources for AWS

[Getting Started](#)

[Training and Certification](#)

[AWS Solutions Portfolio](#)

[Architecture Center](#)

[Product and Technical FAQs](#)

[Analyst Reports](#)

[AWS Partners](#)

### Developers on AWS

[Developer Center](#)

[SDKs & Tools](#)

[.NET on AWS](#)

[Python on AWS](#)

[Java on AWS](#)

[PHP on AWS](#)

[JavaScript on AWS](#)

### Help

[Contact Us](#)

[File a Support Ticket](#)

[Knowledge Center](#)

[AWS re:Post](#)

[AWS Support Overview](#)

[Legal](#)

[AWS Careers](#)

[Create an AWS Account](#)



Amazon is an Equal Opportunity Employer: *Minority / Women / Disability / Veteran / Gender Identity / Sexual Orientation / Age.*

#### Language

[عربي](#) |

[Bahasa Indonesia](#) |

[Deutsch](#) |

[English](#) |

[Español](#) |

[Français](#) |

[Italiano](#) |

[Português](#) |

[Tiếng Việt](#) |

[Türkçe](#) |

[Русский](#) |

[ไทย](#) |

[日本語](#) |

[한국어](#) |

[中文 \(简体\)](#) |

[中文 \(繁體\)](#)

[Privacy](#)

|

[Site Terms](#)

|

[Cookie Preferences](#)

|

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.