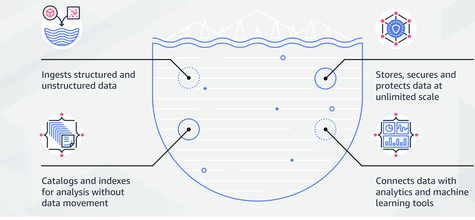
Build your Data Lake on Amazon S3

What is a data lake?

A data lake is a centralized repository that allows you to migrate and store all structured and unstructured data at unlimited scale, as well as gain insights through big data analytics and machine learning tools.

Key components of a data lake



Unprecedented data growth from new and diverse sources

of data worldwide generated **in the last 5 years**

163 ZB

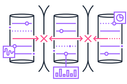
Data creation will grow to 163 zettabytes (ZB) by 2025

but most companies only analyze 12% of their data

Why?

Challenges to gaining the full value of your data

Disconnected data silos



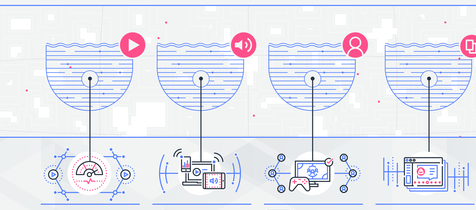
Growing on-premises costs

Incompatible data formats

Break down data silos. **Build your data lake on Amazon S3.**

Amazon S3 is the largest and most performant storage service for structured and unstructured data, allowing you to cost-effectively build and scale a data lake of any size in a secure environment.

10,000+ data lakes on Amazon S3



NETFLIX

delivers billions of hours of content and runs analytics on an S3 data lake

SONOS

1 billion events per week from connected devices

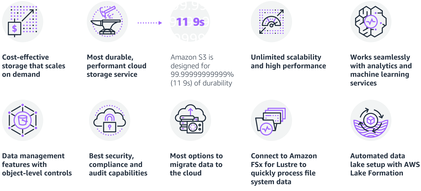
EPIC GAMES

analyzes satisfaction of 125 million players to drive engagement

Georgia-Pacific

analyzes equipment to predict failures to save millions

Why Amazon S3 is the best place to build a data lake



Cost-effective storage that scales on demand

Most durable, performant cloud storage service

Amazon S3 is designed for 99.999999999% (11 9s) of durability

Unlimited scalability and high performance

Works seamlessly with analytics and machine learning services

Data management features with object-level controls

Best security, compliance and audit capabilities

Most options to migrate data to the cloud

Connect to Amazon FSx for Lustre to quickly process file system data

Automated data lake setup with AWS Lake Formation

Amazon S3 data lake lifecycle

A data lake built on Amazon S3 lets you store everything in one place, dive into your data with flexible access, future-proof your storage, and connect to powerful insights.

Ingest and store data

- Migrate data from a variety of data sources
- Real-time data movement
- Remove siloes with one data lake for structured and unstructured data
- Unmatched scale, durability, security, and performance

Catalog and transform data

- Know your data with better management and higher quality data
- AWS Glue crawls, catalogs, and indexes data for searchability
- AWS Glue automates the effort in building, maintaining and running ETL jobs

Analyze

Run AWS analytics and machine learning services to gain insights

- Amazon Athena	- Amazon SageMaker	- Amazon FSx for Lustre
- Amazon Redshift	- Amazon Rekognition	- Amazon EMR

Extract value from data

- Improve customer interactions
- Guide R&D innovation choices
- Maximize operational efficiencies

Amazon S3 is the storage service of choice to build a data lake.

Get started building your data lake on Amazon S3

https://aws.amazon.com/products/storage/data-lake-storage/