

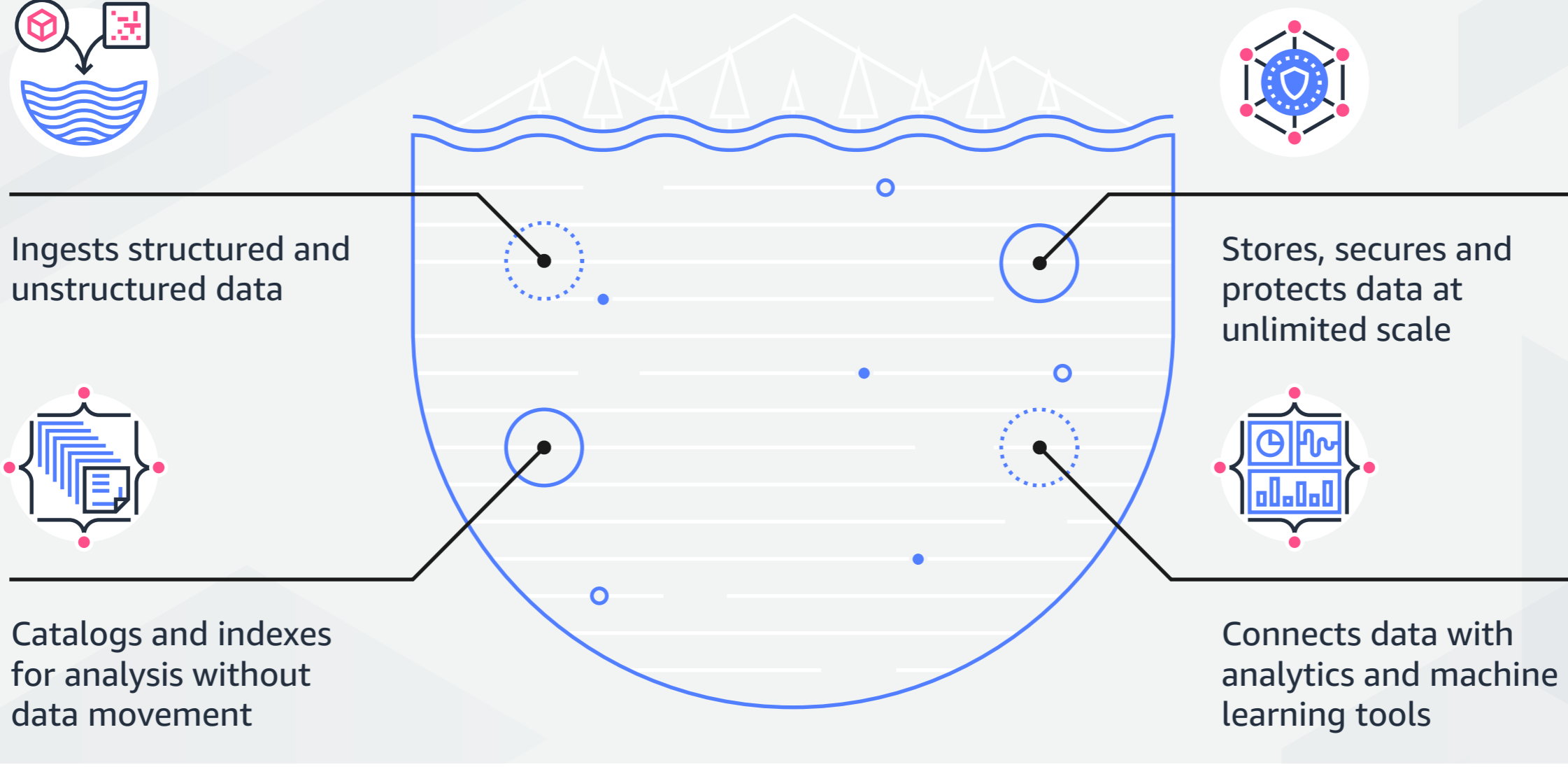
# Build your Data Lake on Amazon S3



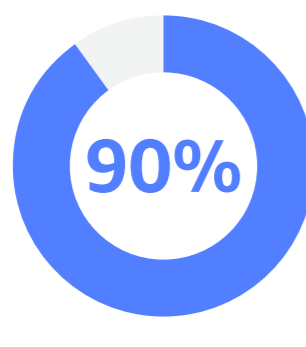
## What is a data lake?

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

12%

but most companies only analyze 12% of their data

### Why?

## Challenges to gaining the full value of your data

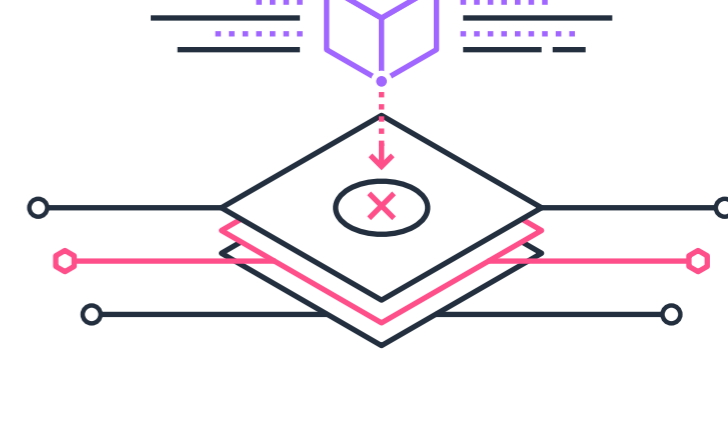
### Disconnected data silos



### 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

The infographic features four data lake examples, each with a play button icon above a lake graphic and an icon below representing the company's data use case:

- 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
- GP Georgia-Pacific:** analyzes equipment to predict failures to save millions

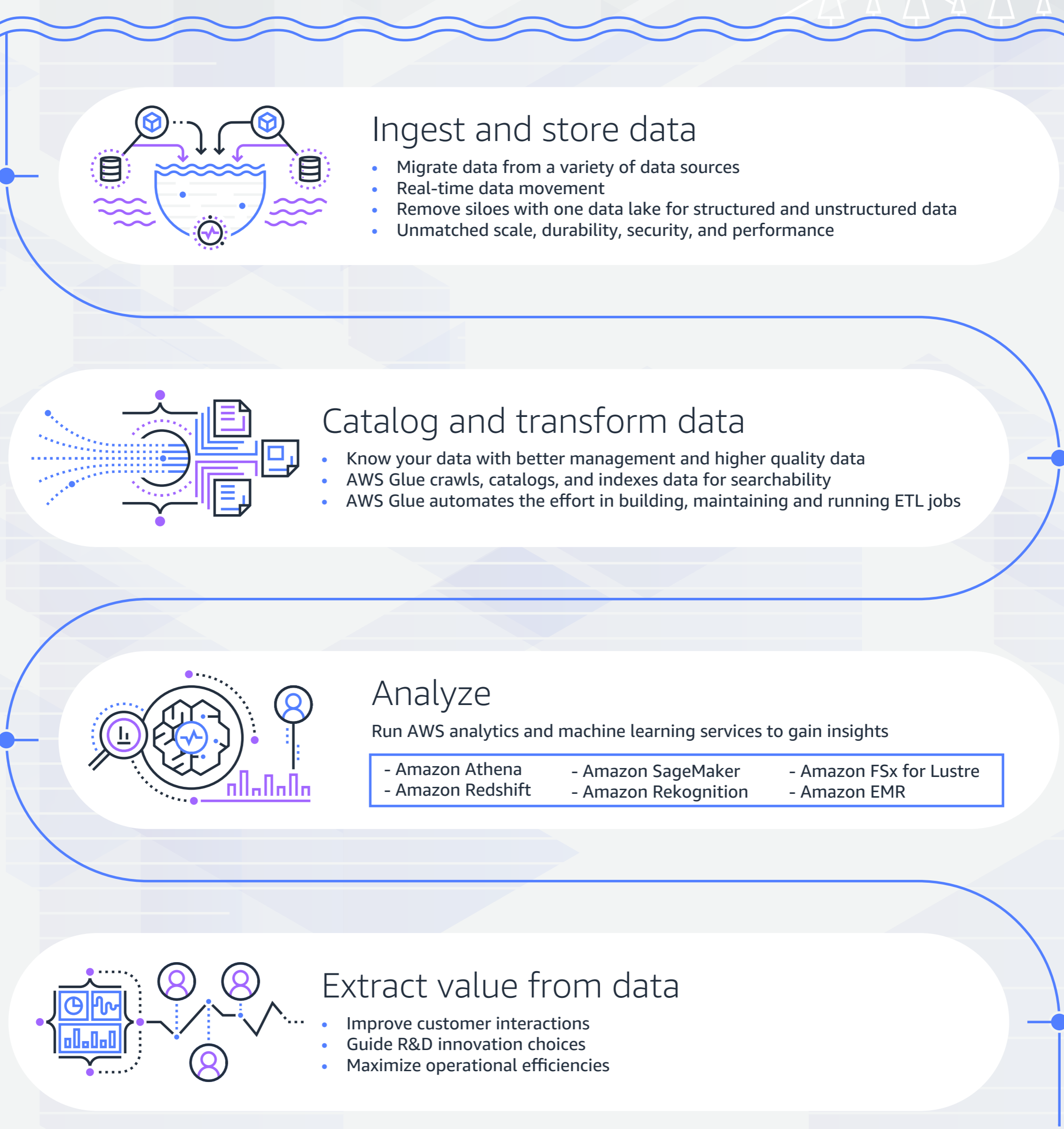
## Why Amazon S3 is the best place to build a data lake

The infographic lists ten benefits of Amazon S3, each with an icon:

- Cost-effective storage that scales on demand:** Represented by a dollar sign and a graph.
- Most durable, performant cloud storage service:** Represented by a scale of justice.
- Amazon S3 is designed for 99.9999999999% (11 9s) of durability:** Represented by a percentage sign and a cloud.
- Unlimited scalability and high performance:** Represented by a magnifying glass and a graph.
- Works seamlessly with analytics and machine learning services:** Represented by a gear and a brain.
- Data management features with object-level controls:** Represented by a shield and a document.
- Best security, compliance and audit capabilities:** Represented by a padlock and a document.
- Most options to migrate data to the cloud:** Represented by a cloud and a server rack.
- Connect to Amazon FSx for Lustre to quickly process file system data:** Represented by a document and a server rack.
- Automated data lake setup with AWS Lake Formation:** Represented by a gear and a document.

## 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.



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/>