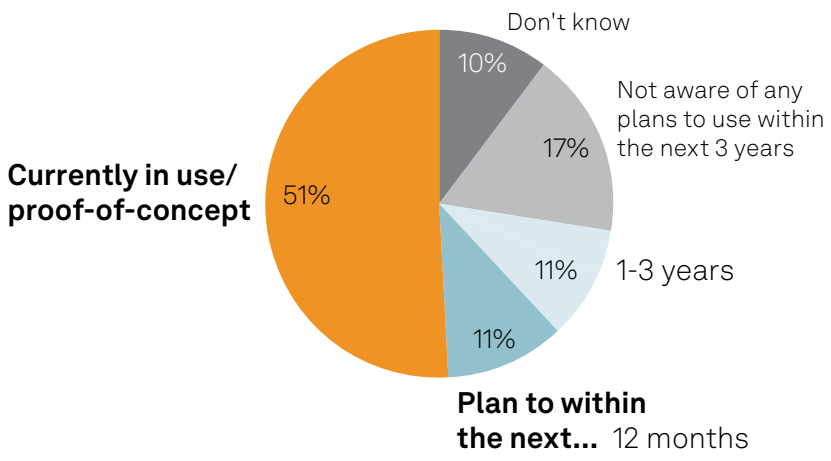
S&P Global
Market Intelligence

Data Lakes

Data lakes are a common data environment

More than half of surveyed enterprises have a data lake, with another **22%** citing plans for a data lake within 36 months.



Q. Which of the following best describes your organization's adoption of data lake or data lakehouse environment(s)?
Base: All respondents (n=1,357).
Source: 451 Research's Voice of the Enterprise: Data & Analytics, Data Platforms & Real-Time Analytics 2023.

This prevalence makes sense because data lakes **provide optionality**

Enterprises can run many types of workloads: operational reporting, business analytics, machine learning model training, data engineering, data cleansing and much more.

66% Analytics

56% Operational applications

53% Data science

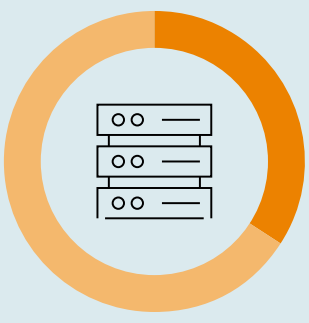
45% Migration or staging

Q. What application use cases run (or will run) on your primary data lake environment? Please select all that apply.
Base: Respondents whose organization currently has a data lake in use, in a proof-of-concept or pilot stage, or plan to use within the next 12 months (n=294).
Source: 451 Research's Voice of the Enterprise: Data & Analytics, Data Platforms, 2022.

But the real power of a data lake lies in **object storage**

Object storage encompasses a sophisticated set of tools and functions that enable a variety of workload types. Data lakes today are powered by cloud object storage, incorporating features such as data durability, elasticity and massive scalability.

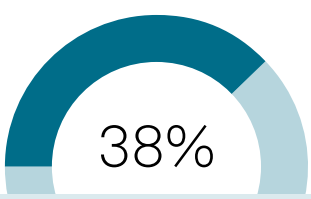
66% of respondents say **public cloud** is the object storage environment that is or will be used for their primary data lake environment

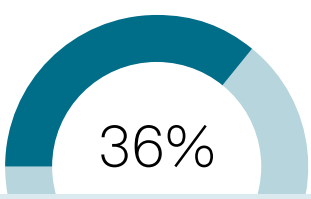


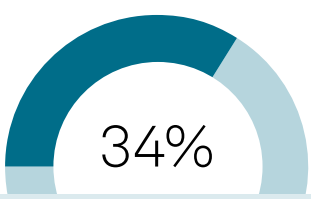
34% in an on-premises data center

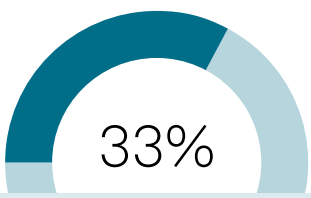
Q. In which location is the object storage environment that is used (or will be used) for your primary data lake environment?
Base: Respondents whose organization currently has a data lake in use, in a proof-of-concept or pilot stage, or plan to use within the next 12 months (n=292).
Source: 451 Research's Voice of the Enterprise: Data & Analytics, Data Platforms, 2022.

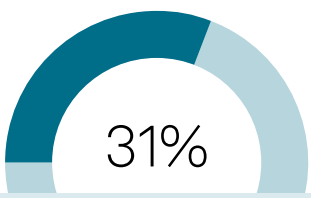
The widespread and growing use of data lakes suggests that enterprises broadly recognize the **significant benefits provided by data lakes**


38%
Improving/developing products and services


36%
Improving business agility


34%
Improving regulatory compliance

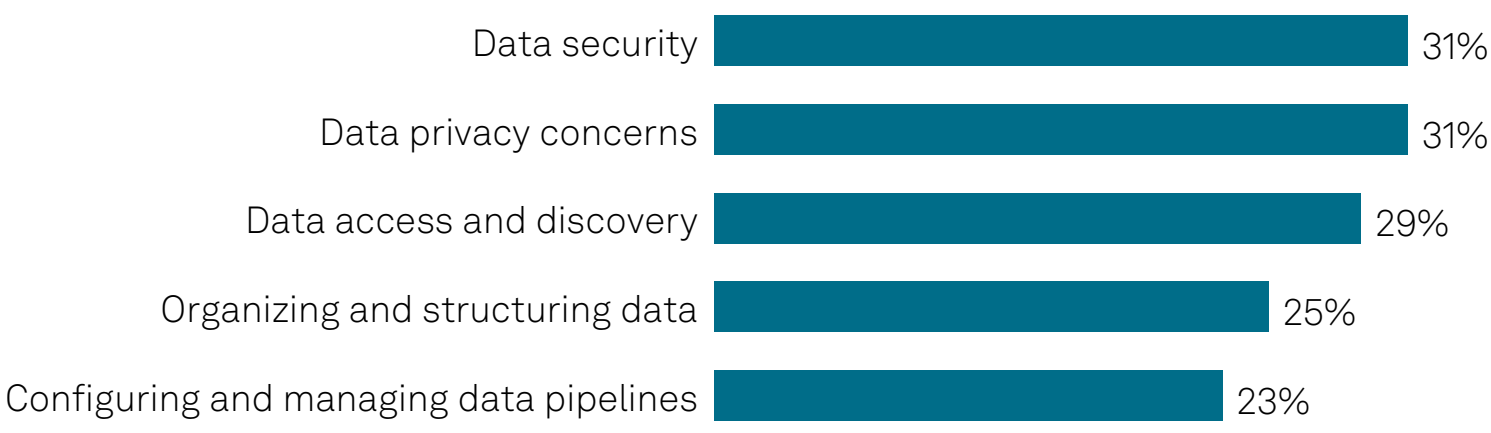

33%
Increasing competitive advantage/responding to competitive threats


31%
Developing data monetization services

Q. What are the most significant benefits your organization expects from your data lake environment? Please select all that apply.
Base: Respondents whose organizations primary approach is to keep existing data lake deployment(s) (n=190).
Source: 451 Research's Voice of the Enterprise: Data & Analytics, Data Platforms & Real-Time Analytics 2023.

Key challenges remain

For example, businesses must maintain security and privacy while also supporting access and discovery of data to drive business goals. Efforts to organize and structure the data, as well as configure and manage data pipelines, are also notable challenges that enterprises have expressed.



Q. And what are the most significant challenges your organization faces in generating insight from your data lake environment? Please select all that apply.
Base: Respondents whose organizations primary approach is to keep existing data lake deployment(s) (n=191).
Source: 451 Research's Voice of the Enterprise: Data & Analytics, Data Platforms & Real-Time Analytics 2023.