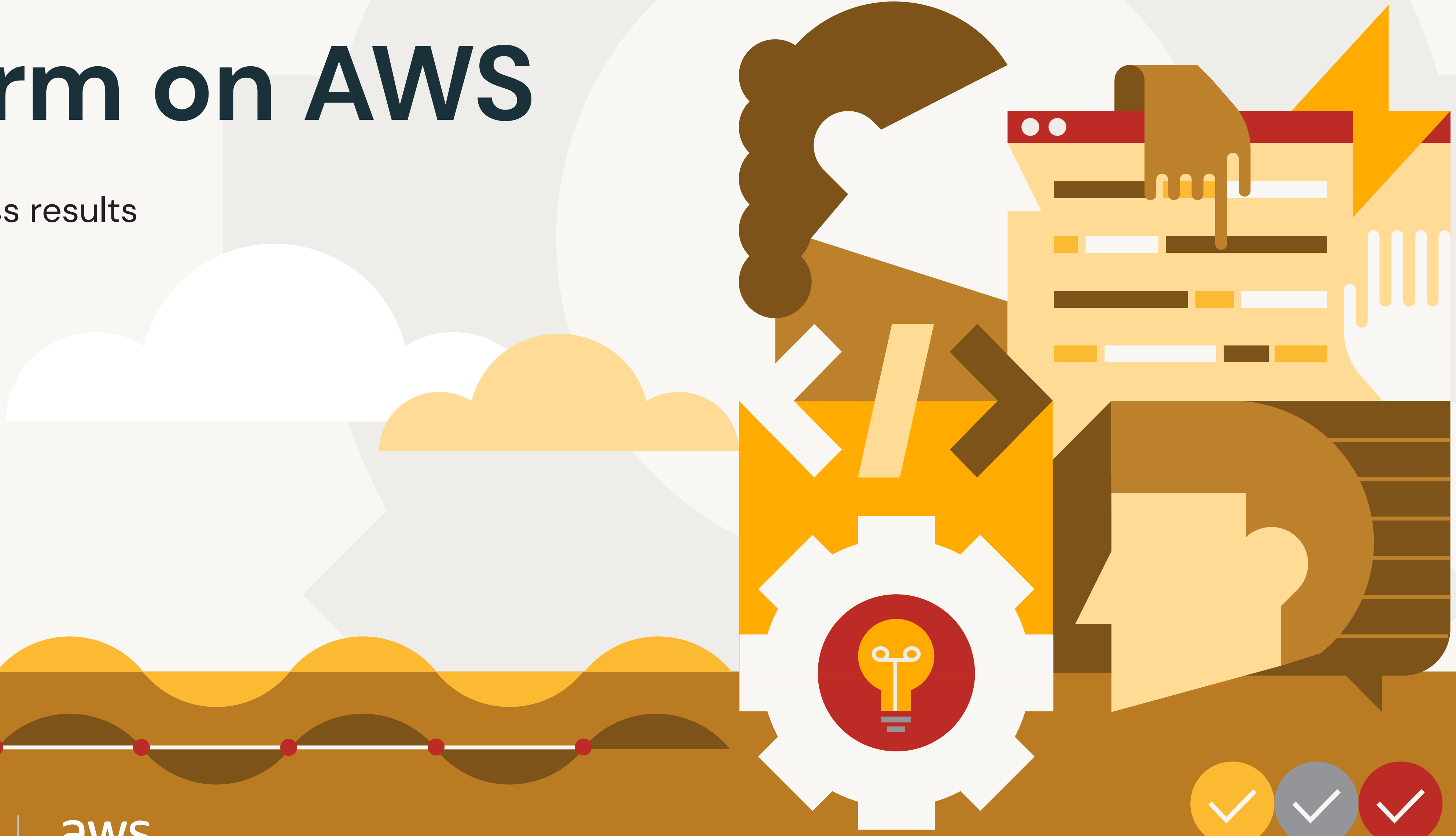
eBook

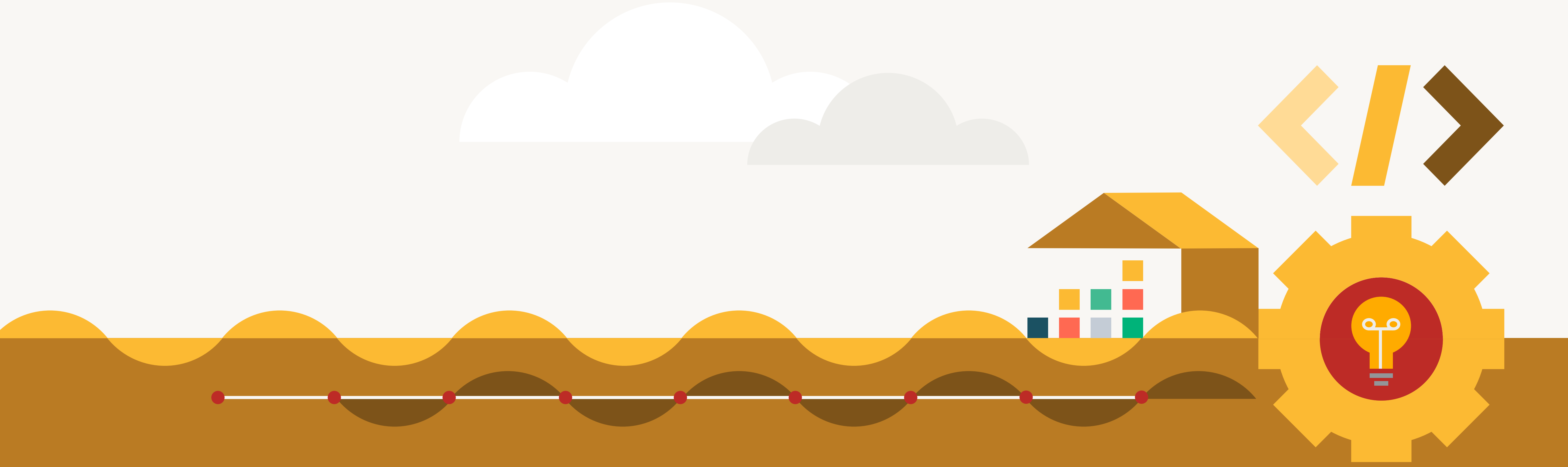
Building a Modern Data Platform on AWS

Accelerate business results with the lakehouse



databricks | aws

Contents



Introduction

Databricks commissioned an independent research firm to survey a random sample of

251

data engineering leaders in the U.S. and Europe

A modern data and analytics platform is essential to your digital transformation strategy. Yet many organizations feel overwhelmed by the complexity of their existing infrastructure, complex extract, transform and load (ETL) data pipelines, and the DevOps requirements of their legacy on-premises data warehouse and big data deployments.

Databricks has helped thousands of companies, including enterprises, migrate to the cloud to lower their costs and improve productivity. The Databricks Lakehouse Platform on AWS enables you to store and manage all your data on a simple, open lakehouse platform that combines the best of data warehouses and data lakes to unify all your data and analytics workloads.

To better understand the key challenges faced by organizations and how the lakehouse paradigm helps customers solve those challenges, Databricks commissioned an independent research firm to survey a random sample of 251 data engineering leaders in the U.S. and Europe about how they're using analytics, AI, data lakes and data lakehouses to drive innovation and growth.[1]

The insights from this study highlight the most successful strategies to address challenges so you can build a successful data platform strategy to drive innovation, increase productivity and accelerate business results.

[1] The margin of error for this study is +/-6% at the 95% confidence level.

The challenges of traditional data platforms

Fast-growing data volumes and higher expectations for data processing are placing more demands on data teams than ever before.

In particular, data engineers face a daunting task of ingesting, cleansing and preparing data for high-value use cases, such as creating new data products and services, enhancing product quality and improving customer service. Delivering high-quality data and meeting service level agreements (SLAs) are essential to enabling data science, machine learning and business intelligence.

Rapid growth of data creates new challenges for data teams. Fifty-two percent of organizations are increasing their budgets to manage an ever-increasing quantity of data. Fifty-two percent are deploying more tech tools, and 41% are bringing on more staff to manage the rising tide of data.

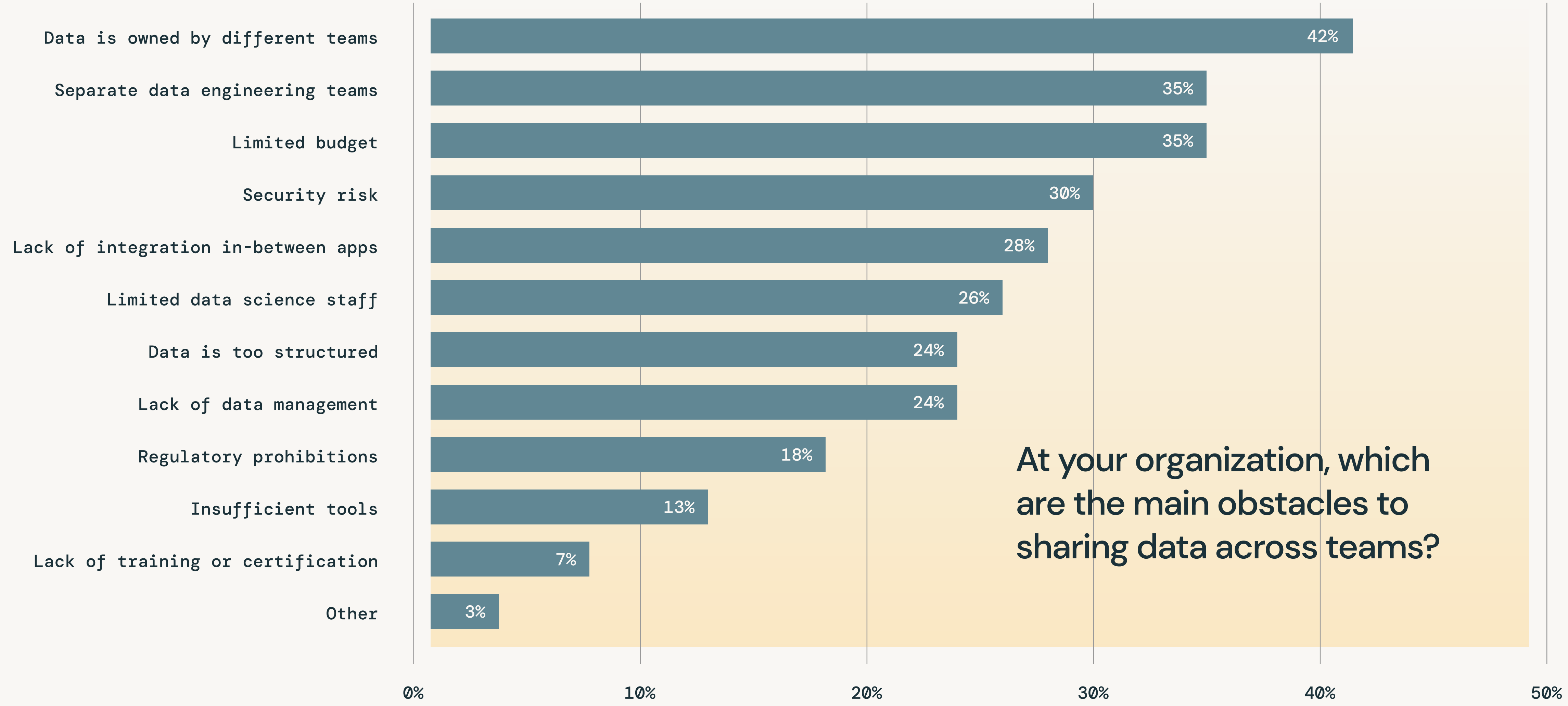
Despite investments in budgets, tech tools and staff, organizations face numerous obstacles to data sharing. Forty-two percent say that data ownership by different teams is the main obstacle to sharing data across teams.

94%

of data engineers say that meeting their SLAs is a top priority, but only 45% are always able to meet them

92%

of organizations have at least doubled the quantity of data they manage in the past 3 years



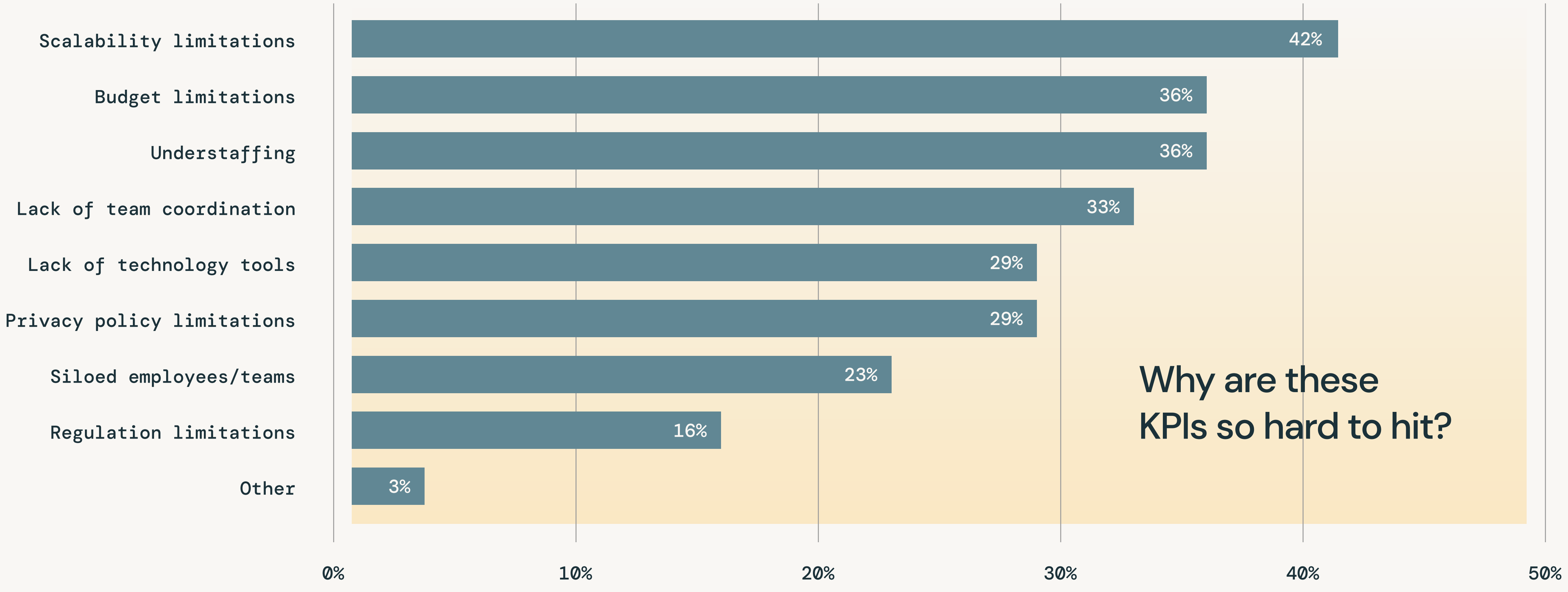
At your organization, which are the main obstacles to sharing data across teams?

83%

of data engineers think meeting their KPIs will get harder over the next 2 years

Key performance indicators (KPIs) are getting harder for data teams to reach. When asked which of their data lake or data warehouse KPIs are hardest to hit, 39% of data engineers say “volume of data” and 34% say “number of data sources.” More than one-fifth say that “security” is their most challenging KPI.

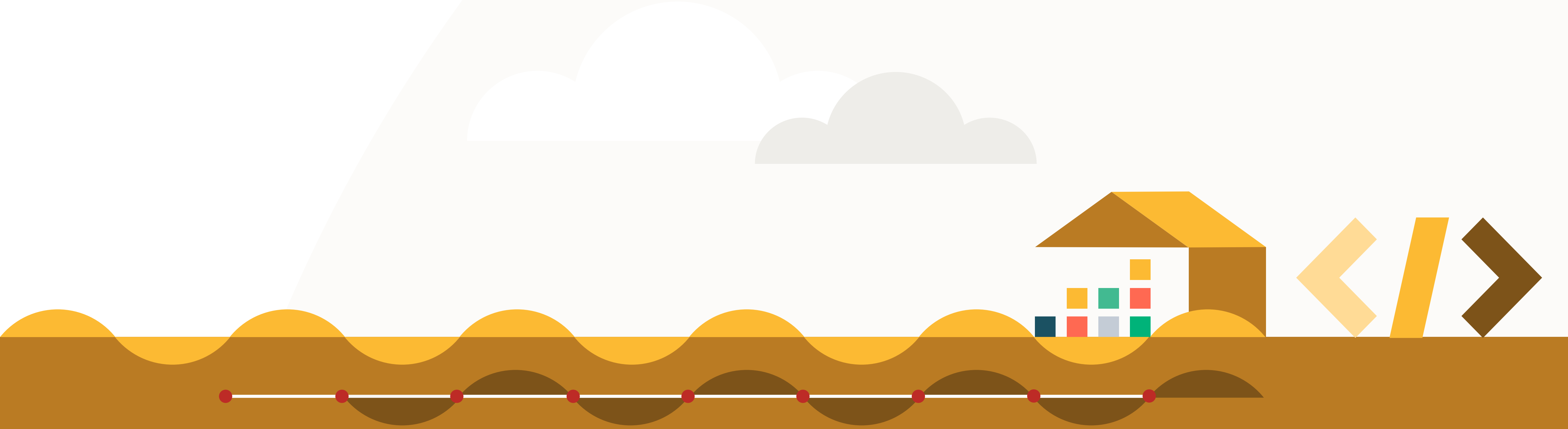
Scalability limitations are the most common reason data teams are struggling to hit their KPIs. And while many organizations are increasing budgets and staffing to address scale issues, 36% of organizations are encountering limits in these areas.



Companies increasingly depend on data to serve both internal users and customers. Ninety-one percent of organizations say they use their data insights to develop successful new products and services, and 78% of organizations say their data lake or data warehouse is a revenue source. Turning new and existing data sources into value is a challenge. Seventy-four percent of organizations struggle to monetize the data they collect.

74%

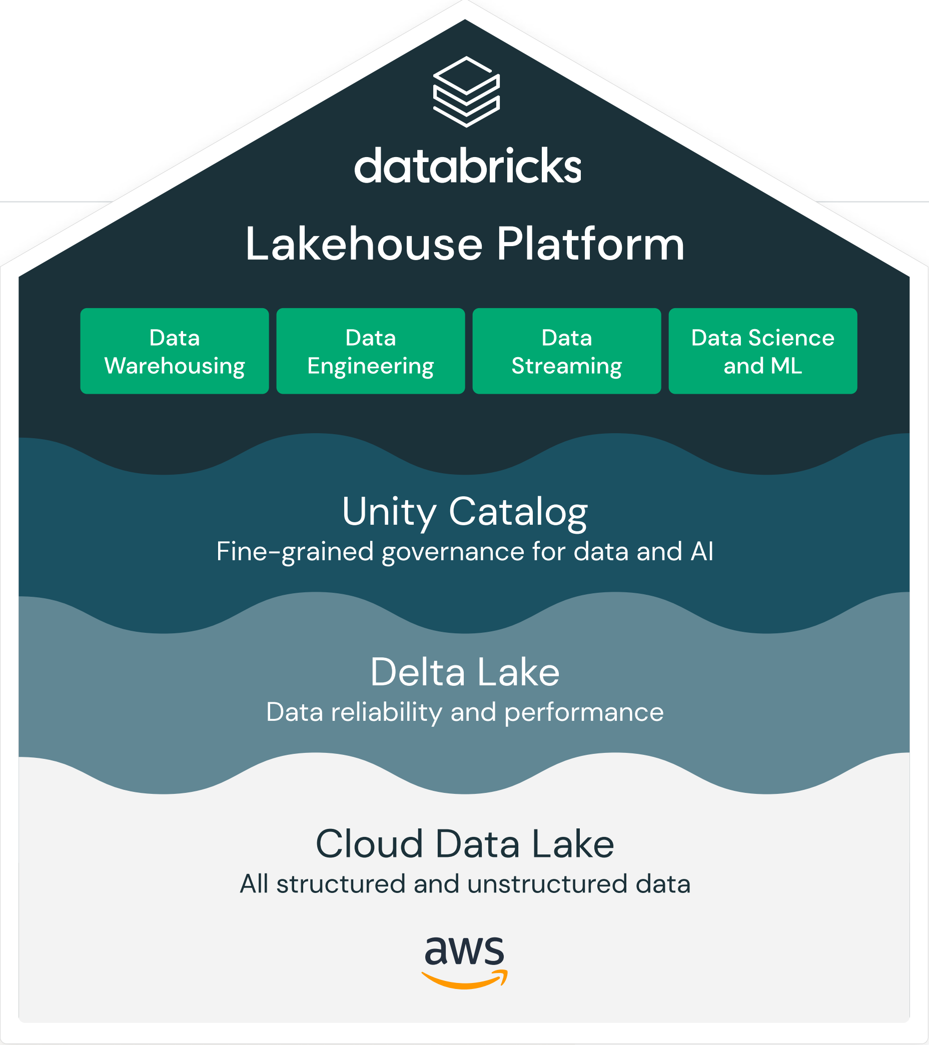
of organizations say they struggle to monetize the data they collect



The data lakehouse — a simpler approach

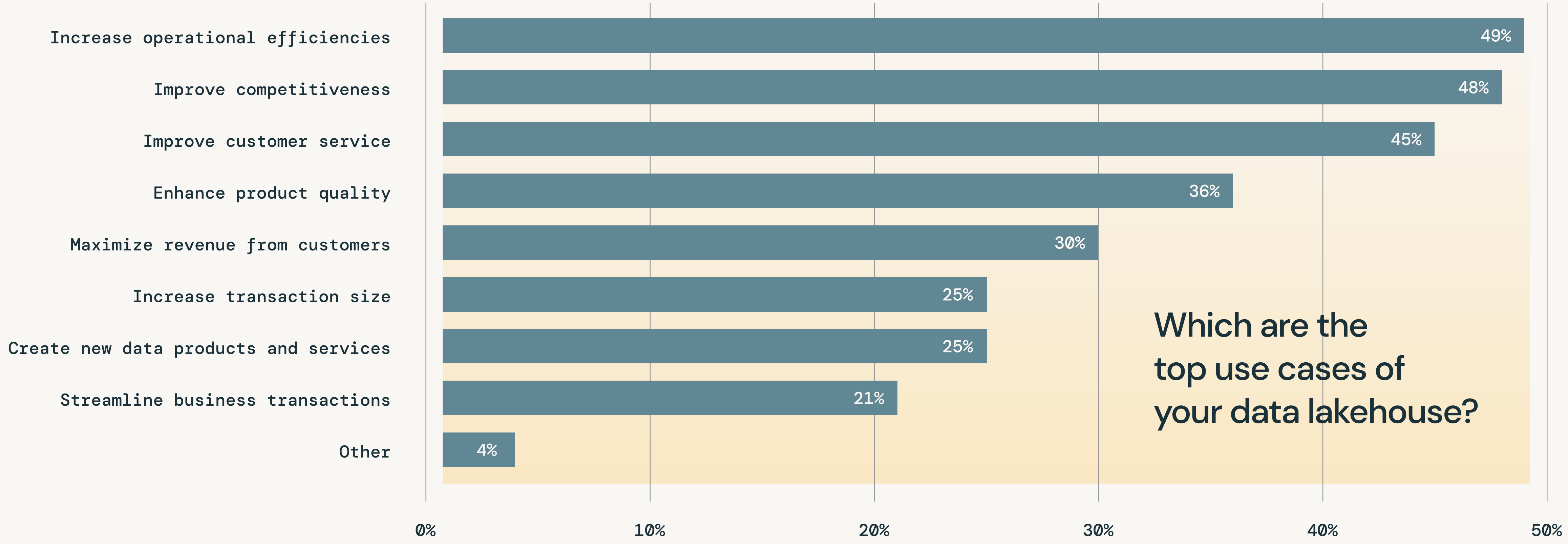
A data lakehouse combines the capabilities of a data warehouse with a data lake. Lakehouse architectures are built on an open, reliable and consistent foundation, supporting a wide variety of structured, semi-structured and unstructured data. This foundation simplifies data sharing and provides a more scalable and efficient approach to analytics, data science, machine learning and business intelligence (BI).

Data teams are already experiencing the benefits of the lakehouse across a variety of data analytics use cases. A data lakehouse helps increase operational efficiencies, improve competitiveness, improve customer service, enhance product quality and more.



97%

of data engineers using a data lakehouse say it simplifies their work



Which are the top use cases of your data lakehouse?

98%

of AWS customers using Databricks achieve their AI ROI goals

A lakehouse architecture helps data teams reach their KPI and return on investment (ROI) goals and turn data into revenue streams. Data engineers who don't use a data lakehouse vs. those who do are nearly twice as likely to say KPIs will get harder over the next two years.

Organizations using the Databricks Lakehouse Platform on AWS are better able to achieve their AI goals. Ninety-eight percent of AWS customers using Databricks achieve their AI ROI goals compared with 70% for organizations that don't use AWS or Databricks.

The lakehouse is enabling organizations to successfully create new data products and revenue streams. Fifty-six percent of organizations without a data lakehouse have turned their data into a revenue stream compared to 88% of organizations with a data lakehouse.

Organizations are adopting the lakehouse to simplify their work, scale ELT data pipelines, improve efficiency, achieve their ROI goals and accelerate new data products and revenue streams. Eighty-eight percent of organizations that don't have a data lakehouse expect to have one within two years.

88%

of organizations that don't have a data lakehouse expect to have one within 2 years

Accelerate results with the Databricks Lakehouse Platform on AWS

The Databricks Lakehouse Platform combines the best elements of data lakes and data warehouses to deliver the reliability, strong governance and performance of data warehouses with the openness, flexibility and machine learning support of data lakes.

This unified approach simplifies your modern data stack by eliminating the data silos that traditionally separate and complicate data engineering, analytics, BI, data science and machine learning. It's built on open source and open standards to maximize flexibility. And, its common approach to data management, security and governance helps you operate more efficiently and innovate faster.

Get started today with the Databricks Lakehouse Platform on AWS. Start your free trial and follow step-by-step training.

Free Trial

Try Databricks on AWS

Free Step-by-Step Training

Start learning

About Databricks

Databricks is the lakehouse company. More than 7,000 organizations worldwide — including Comcast, Condé Nast, H&M and over 50% of the Fortune 500 — rely on the Databricks Lakehouse Platform to unify their data, analytics and AI. Databricks is headquartered in San Francisco, with offices around the globe. Founded by the original creators of Apache Spark™, Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems.

To learn more, follow Databricks on Twitter, LinkedIn and Facebook.

databricks | aws