



# An In-Depth Look at Big Data Trends and Challenges

*Results from a Global Survey by Dimensional Research  
August 2018*



## Global Survey of Business, IT and Big Data Professionals

### OVERVIEW

The realization of value from big data in on-premises environments has lagged expectations, largely due to complexity, cost overruns, steep demands for specialized talent, and need for powerful computing platforms.

As a result, companies are rapidly moving big data processing to the cloud, where powerful, cost-effective self-service platforms and elastic computing are readily available.

The move to the cloud has turned big data into a profit center with use cases centered around artificial intelligence (AI), machine learning (ML), and analytics. But as new big data opportunities emerge, so do new challenges.

Sponsored by Qubole, Dimensional Research launched a survey in June 2018 of 401 data professionals with big data responsibilities in enterprises across the globe.

This report provides insights into big data processing trends, challenges, and solutions across enterprises worldwide.



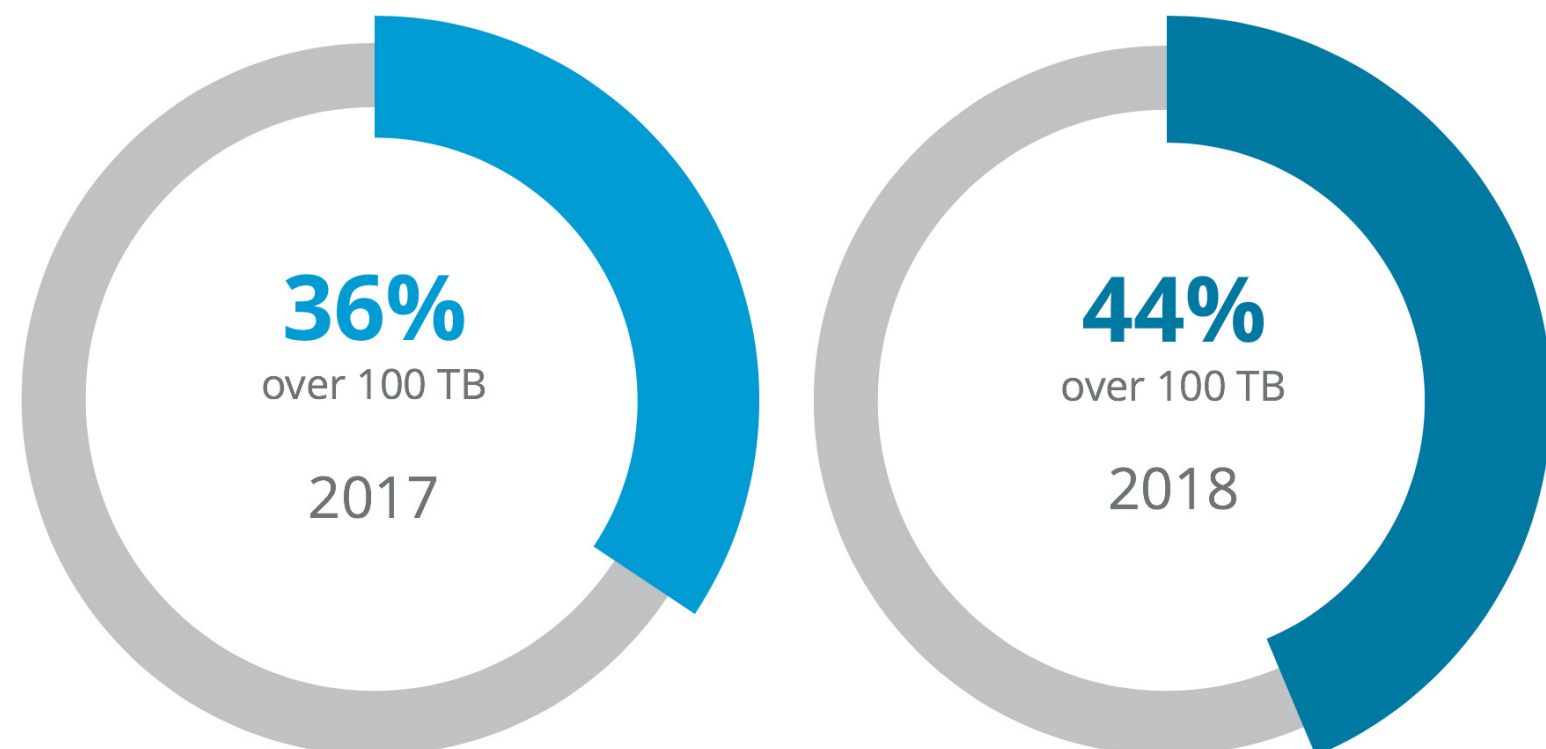
What size  
is your  
data lake?

## Big Data is Skyrocketing in Size

As the digital world provides an ever-rising flood of information, big data lakes in modern organizations are growing to immense sizes and will continue to do so.

Forecasts call for future datasets that far exceed the sizes of today's big data repositories.

### Organizations with Average Data Lake Size Over 100 Terabytes

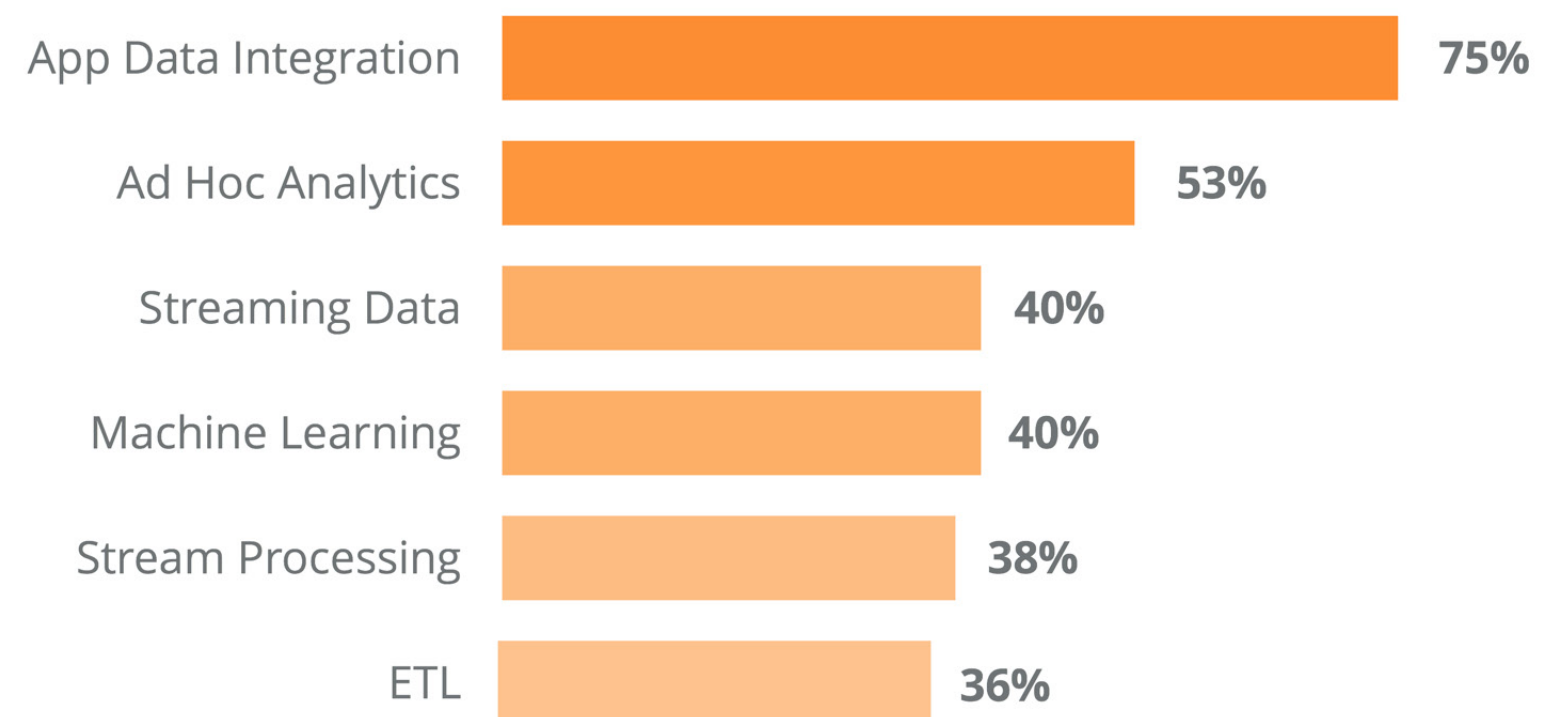


What types  
of big data  
processing do  
you perform?

## Types of Big Data Processing

Big data is a vital component of everyday information processing and a key asset for enterprises. Enterprises process big data continuously for multiple purposes, from analytics to ML, to application integration and more. Big data plays an essential role in delivering new insights, innovating, increasing productivity, and realizing new revenue opportunities.

### Types of Big Data Processing Performed

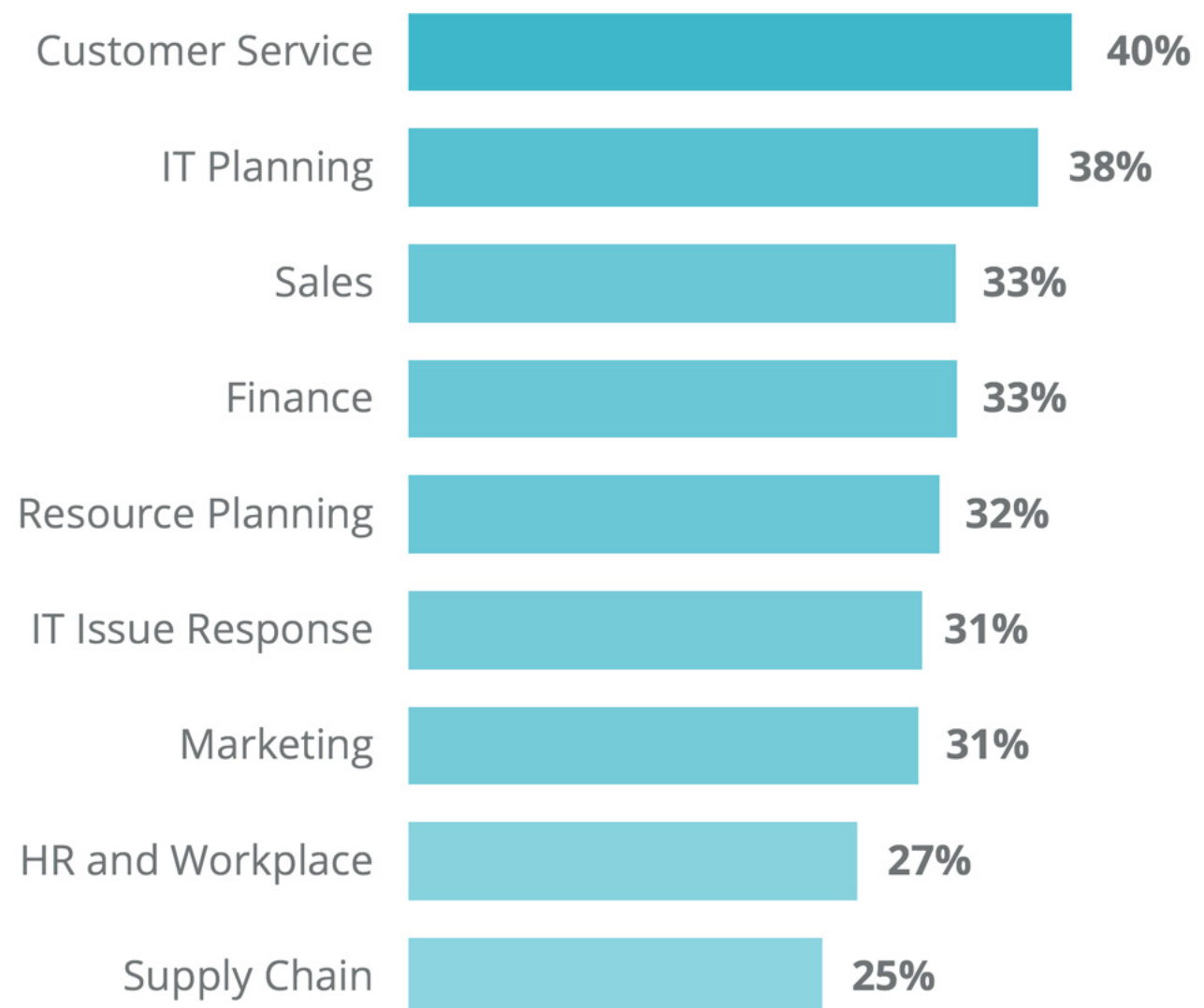


Over the next 12 months, which business processes will receive insights from your big data initiatives?

## Extensive Enterprise Value from Big Data

Big data is being used across a wide and growing spectrum of departments and functions in modern organizations. The insights and intelligence provided by big data translate directly into operational efficiencies and competitive advantage.

### Business Processes Receiving Value from Big Data

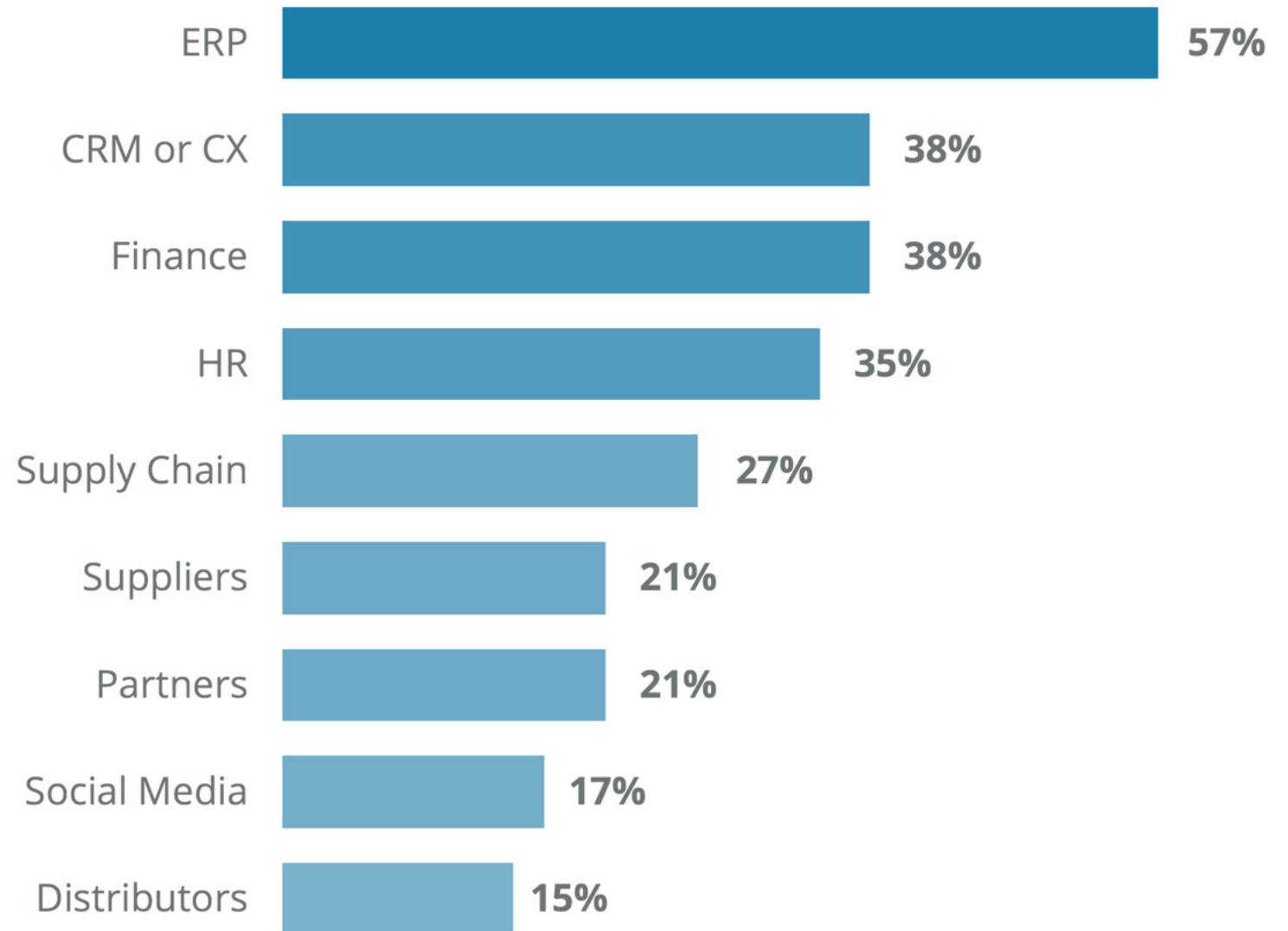


## Sources Feeding the Big Data Firehose

Big data is pouring in from across the extended enterprise, the Internet, and third-party data sources. The staggering volume and diversity of the information mandates the use of frameworks for big data processing.

Which of these data sources does your company use in its big data processing today?

Sources of Big Data

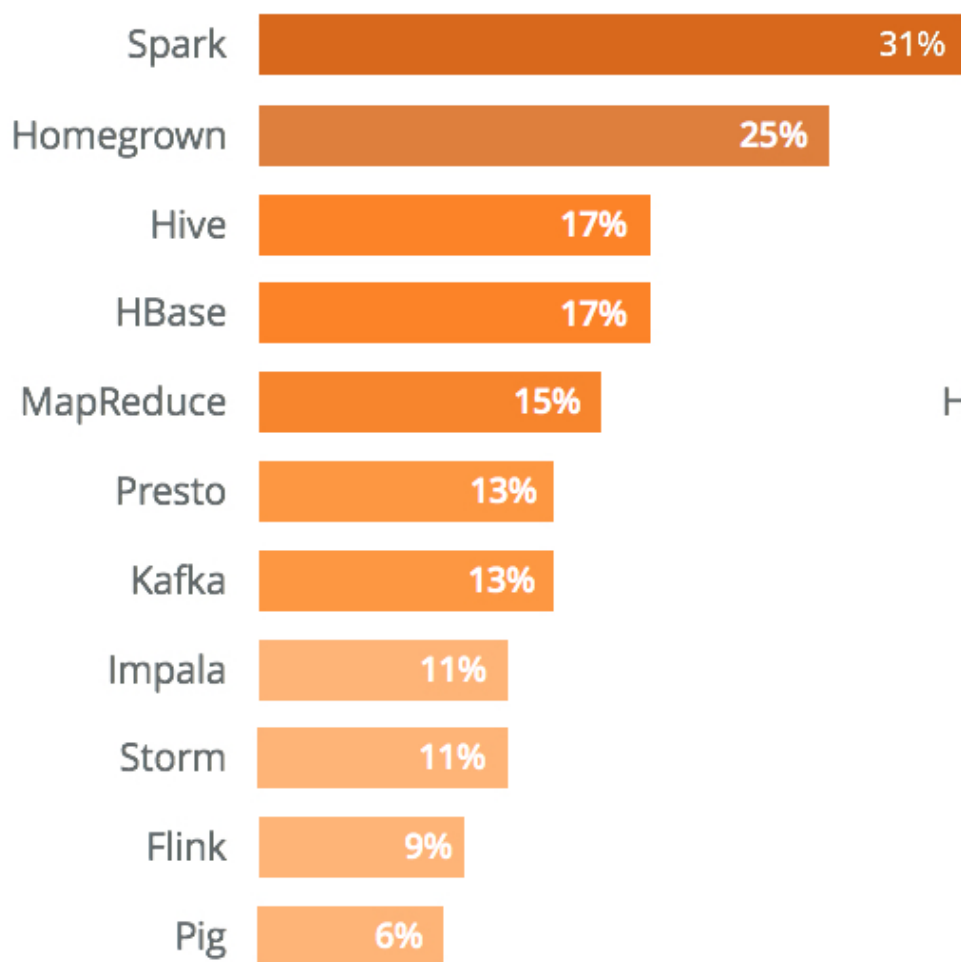


Which big data software framework does your company currently use?

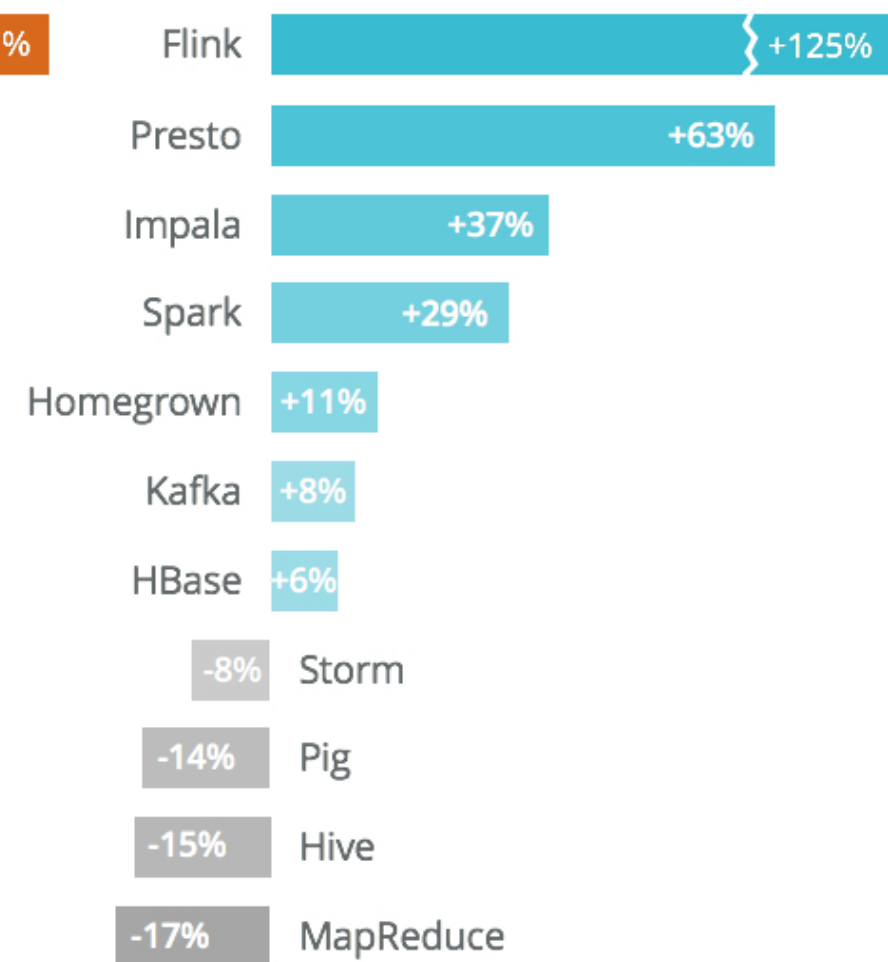
## Big Data Frameworks Are Constantly Evolving

While no single software framework dominates the big data landscape, Apache Spark and Presto are showing impressive gains. Survey data also shows organizations moving from homegrown approaches to open source technologies.

Frameworks in Use in 2018



Percent Change from 2017

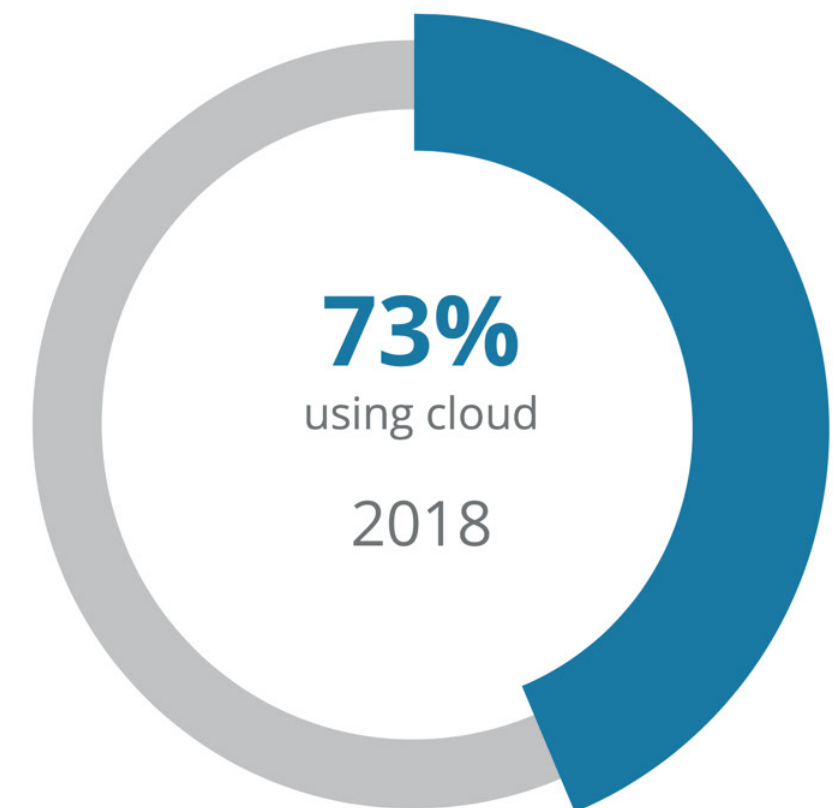
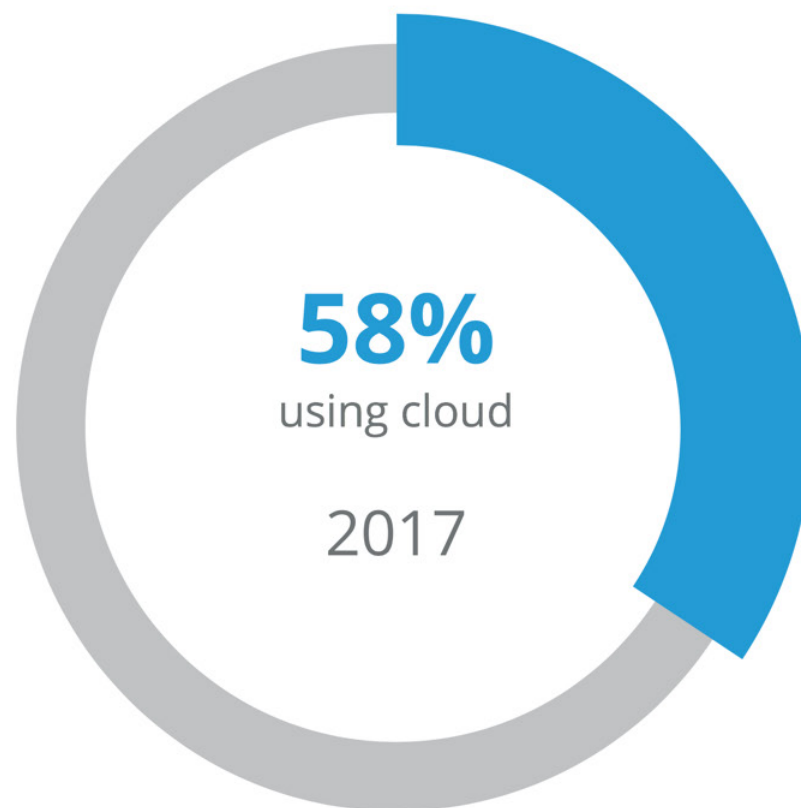


Is your  
big data  
processing  
performed  
in the cloud?

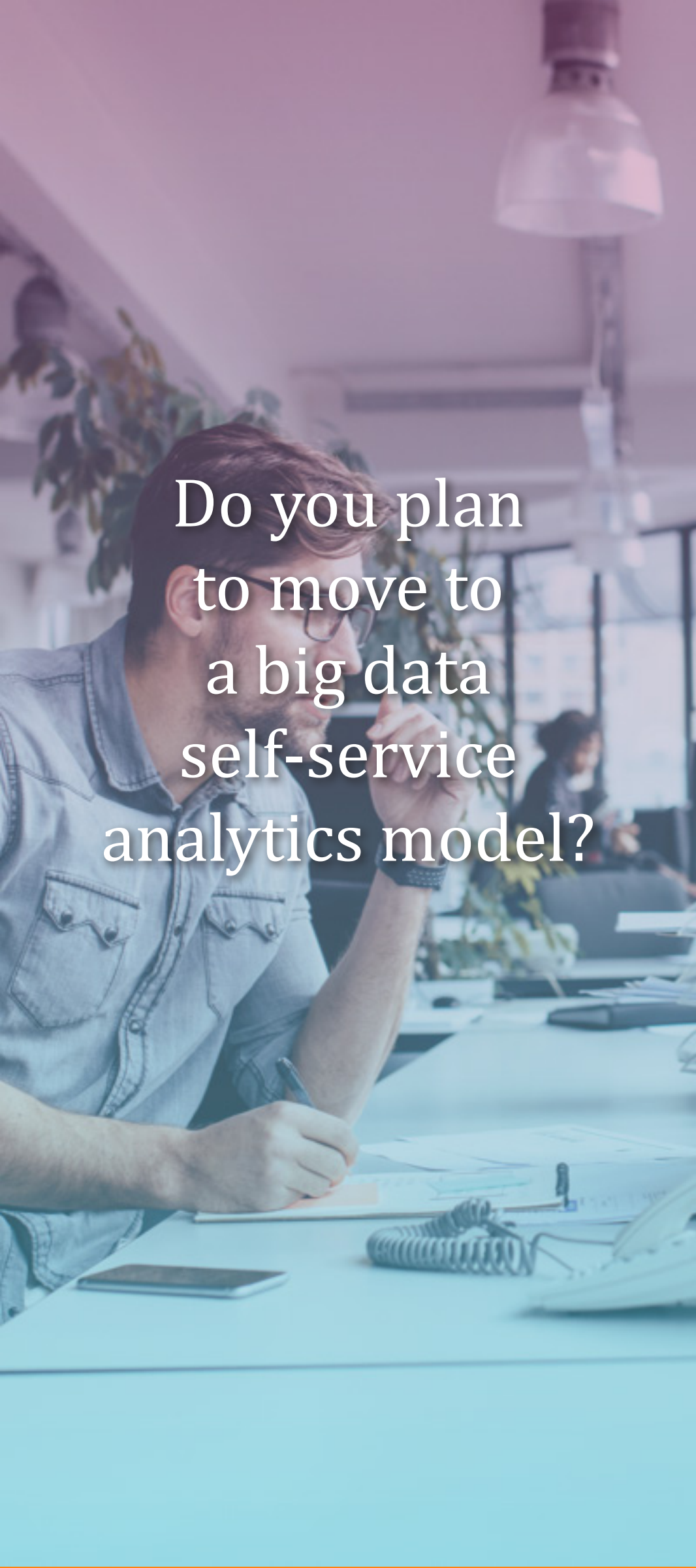
## Big Data Processing is Moving to the Cloud

The overwhelming majority of organizations are moving their big data processing to the cloud to take advantage of its convenience, cost, integration, and performance benefits.

### Organizations Using Cloud for Big Data Processing





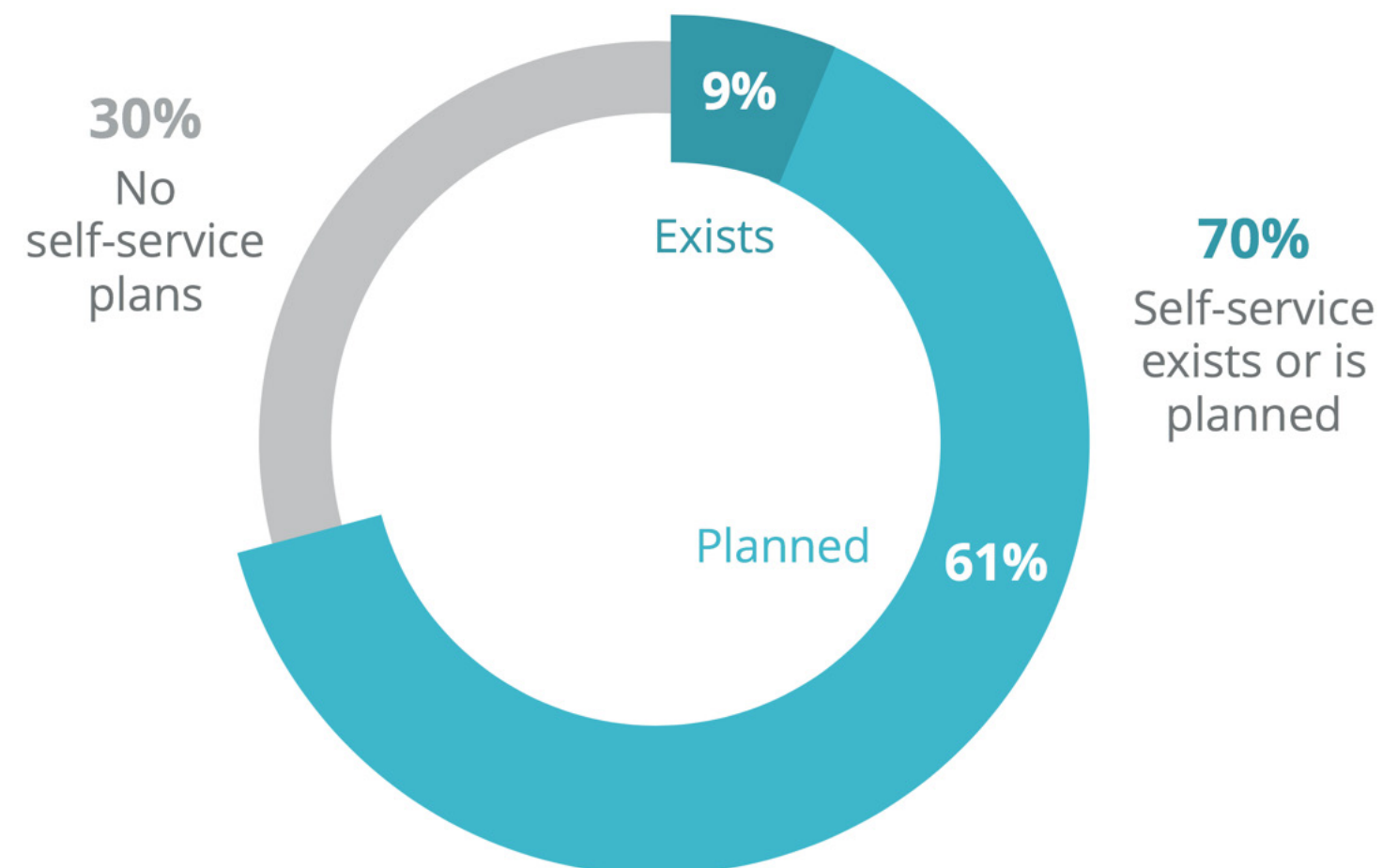


Do you plan  
to move to  
a big data  
self-service  
analytics model?

## The Need for Self-Service Environments

By empowering users to build their own data pipelines to perform analytics and utilize machine learning, big data teams can control staffing costs and raise the productivity of business professionals across the enterprise.

### Plans to Support Self-Service Big Data Analytics

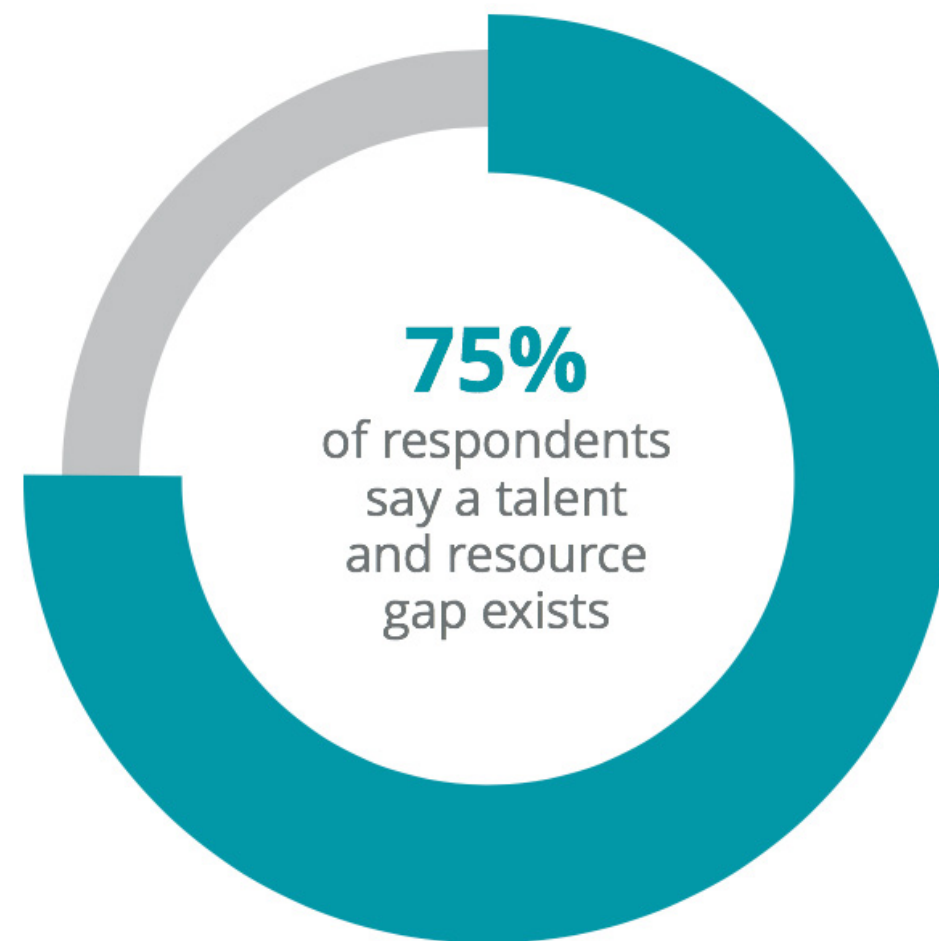


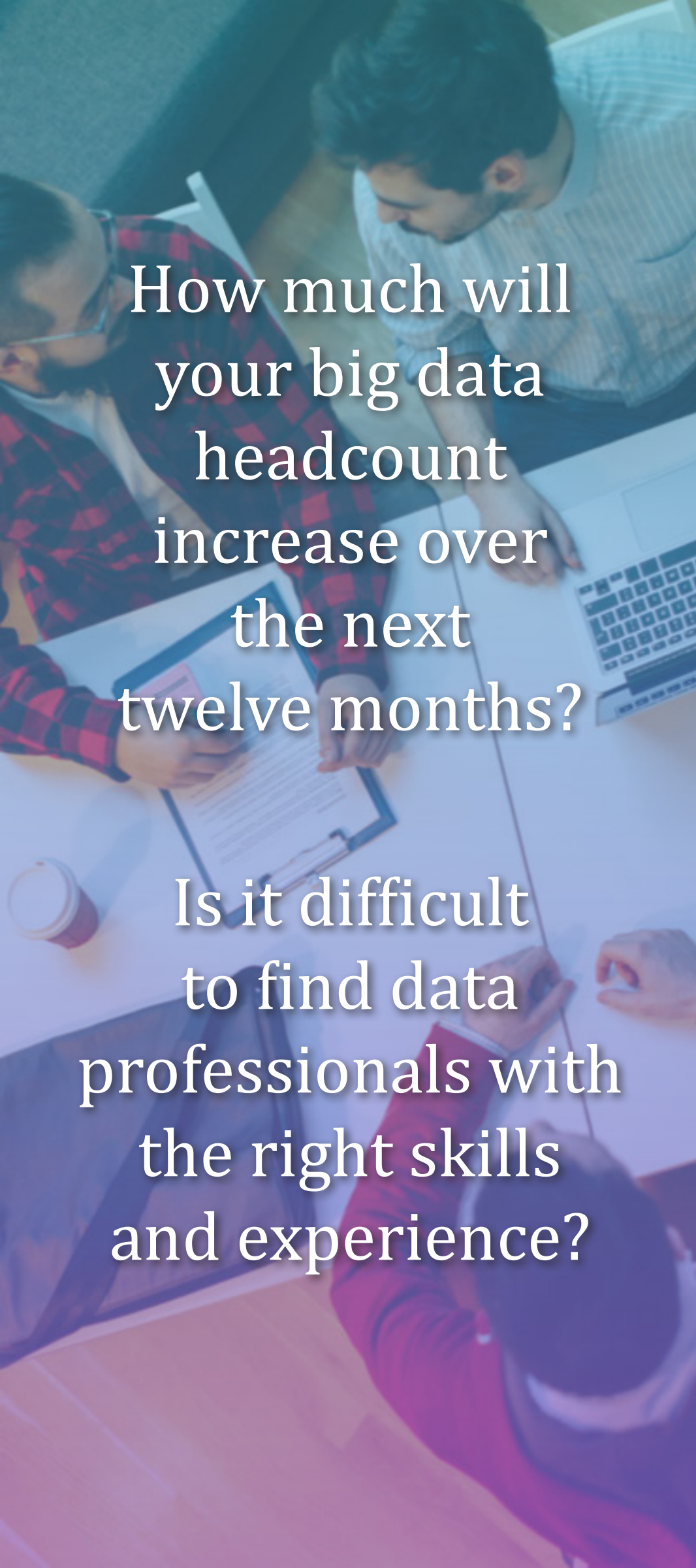
Is there a gap between the potential value of big data and the tools and people dedicated to deliver it?

## Not Enough Dedicated Tools and Resources

Three-quarters of survey respondents noted a sizable gap between the dedicated tools and resources available, and the potential value of their big data projects. This mismatch triggers an inability to transform data into profits and reduce associated operational costs.

Gap Between Big Data Value and Dedicated Resources





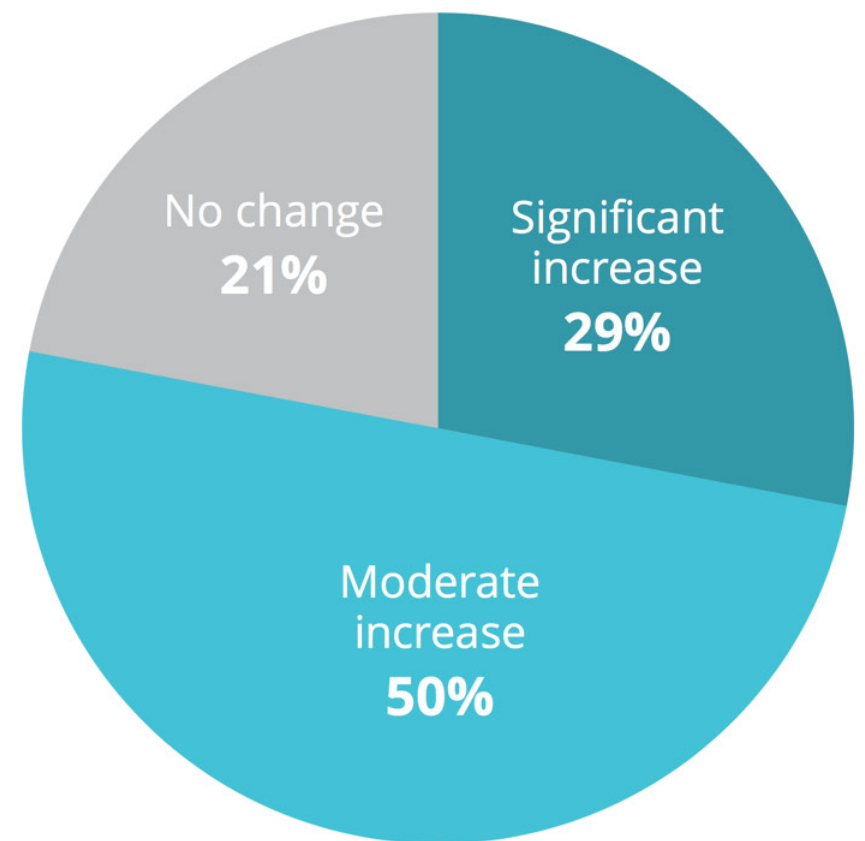
How much will your big data headcount increase over the next twelve months?

Is it difficult to find data professionals with the right skills and experience?

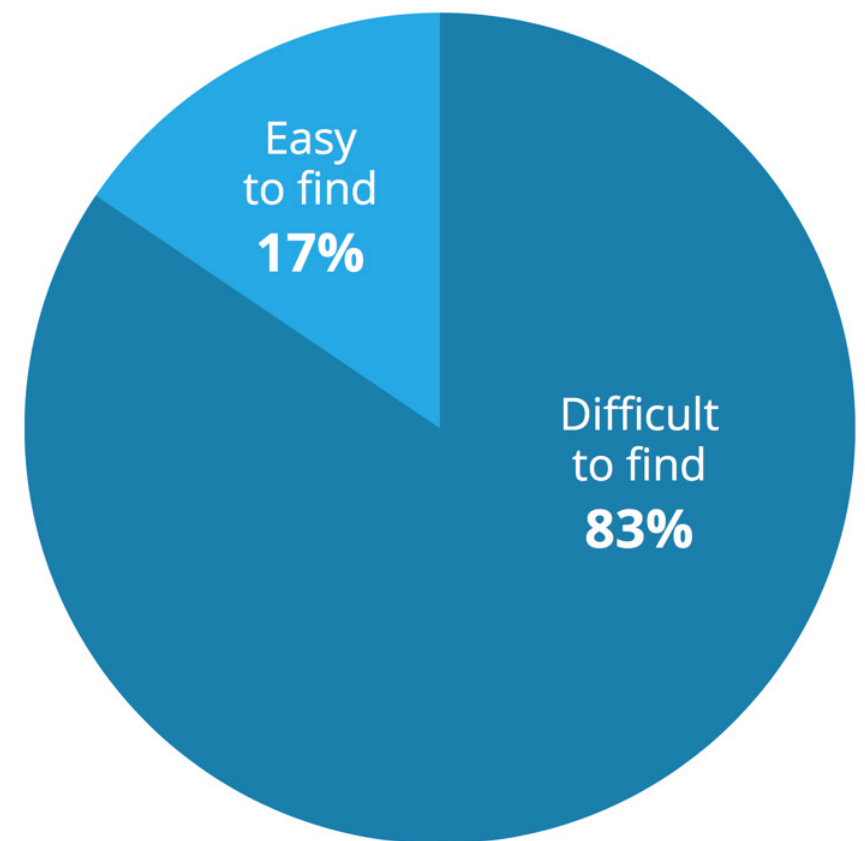
## The Shortage of Big Data Talent is Very Real

As big data becomes a vital part of information strategies, IT organizations would like to grow headcount to meet project demands. But many IT groups have flat budgets and face trouble finding qualified big data talent. The answer lies in maximizing the productivity of existing big data teams.

Big Data Headcount Growth Plans



Ease of Finding Qualified Big Data Talent





What is your  
current ratio  
of big data  
administrators  
to users?

## Big Data Admins Must Serve More Users

Only 40 percent of big data administrators are able to support more than 25 users—a startling number, since today’s flat budgets call for admins to serve more than 100 users. As big data strategies and implementations mature, organizations require platforms that enable much higher admin-to-user ratios.

Users Supported per Big Data Administrator



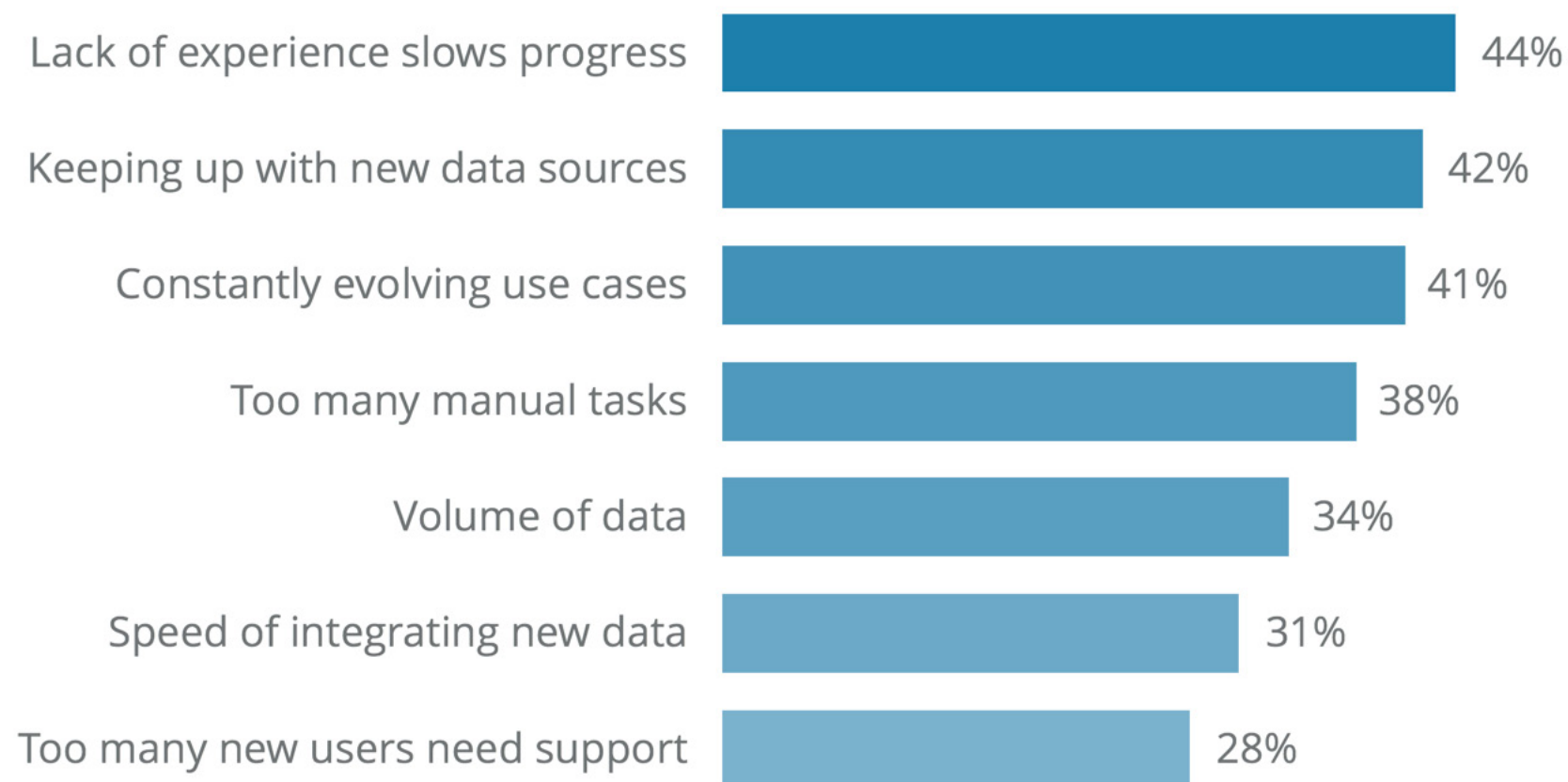


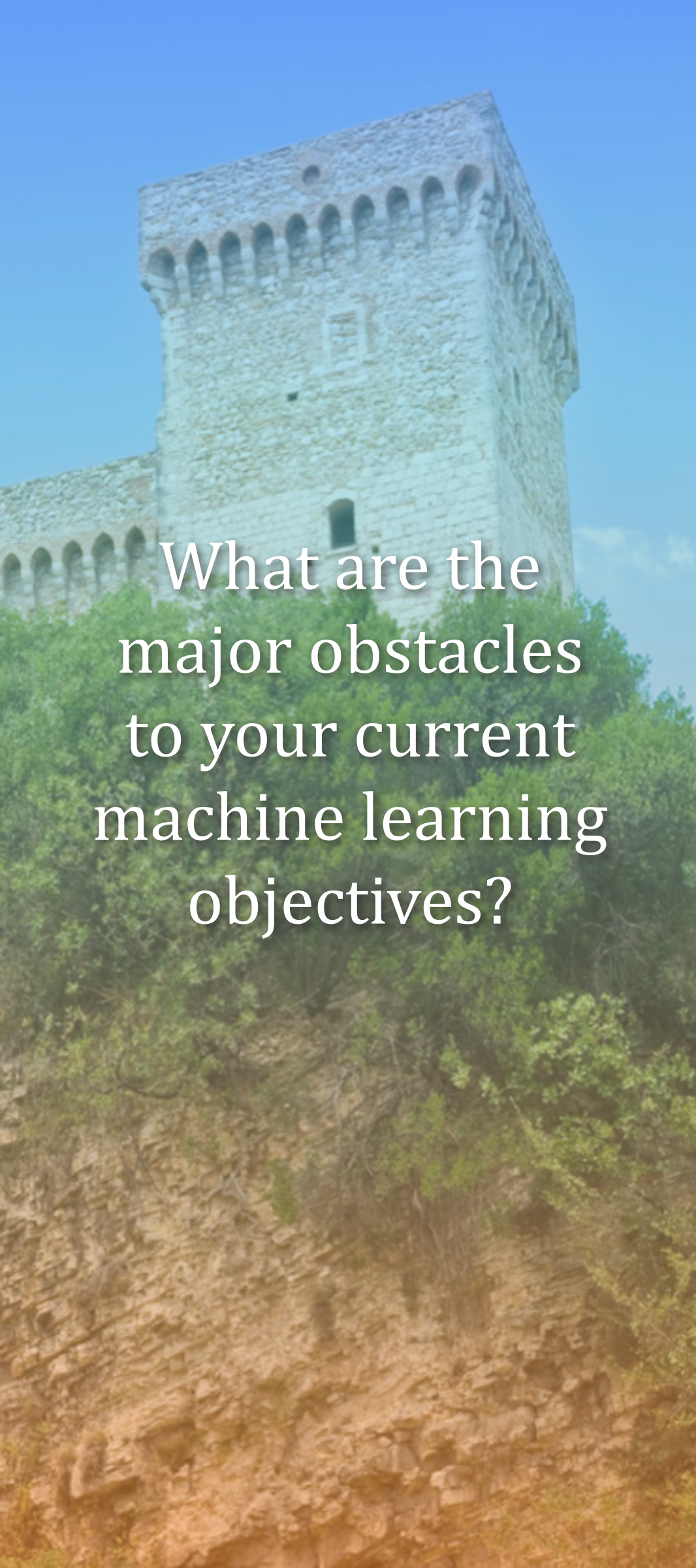
Which of these challenges is your big data team experiencing today?

## Challenges Faced by Big Data Teams

Big data strategies are not without their challenges, which most often are triggered by overwhelming data and computing demands that require special technologies and skill sets. Most big data teams face a variety of issues that need to be addressed with education, headcount, and better tools.

### Most Common Big Data Challenges



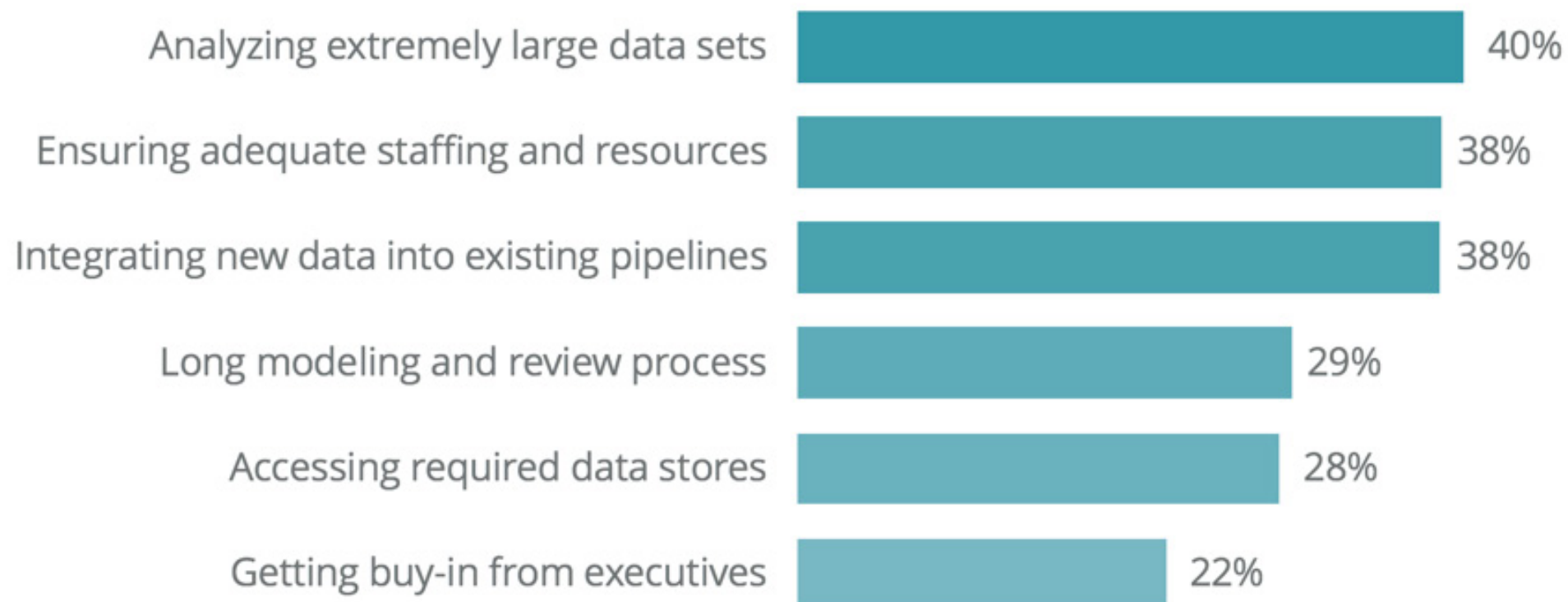


What are the major obstacles to your current machine learning objectives?

## Machine Learning Presents Its Own Obstacles

Machine learning is being used in diverse initiatives that include crucial security, maintenance, customer-care, and lead-generation applications. But like most data-intensive solutions, machine learning presents a variety of implementation challenges.

### Most Common Machine Learning Obstacles



What is the top complaint you hear from executives and stakeholders about big data initiatives?

## Concerns About Big Data Time-to-Value

The most common complaints cited by executives and stakeholders focus on the time and costs required to derive real business value from big data initiatives. These comments illuminate the need for smarter, more productive platforms, tools, and technologists.

### Common Complaints About Big Data Initiatives





Do you need  
a faster, easier  
way to reap  
the rewards  
of big data  
in your  
organization?

## The Cloud-Native, Data Platform for Artificial Intelligence, Machine Learning and Analytics

More than 200 organizations worldwide use Qubole to process over an exabyte of data per month for AI, machine learning, and analytics. Qubole's cloud-native platform delivers:

- Fastest time to value from AI, ML, and analytics
- End-to-end data processing on a single, shared platform
- Self-service capabilities to support 10 times more users and data per administrator
- 50 percent lower cloud data processing costs than alternatives

Since the company's inception, Qubole has continued to push the limits of possibility in big data and cloud computing.

To learn more, visit [qubole.com](https://qubole.com)

