

# Big Data & Analytics Maturity 2020 Survey Results

Survey reveals most enterprises are choosing a multi-cloud strategy and that data virtualization and data governance are top priorities





## INTRODUCTION

In collaboration with Cloudera and ODPi.org (a project of the Linux Foundation), we've collected responses from over 150 big data and analytics leaders in this year's survey. Responses came back from across industries and around the globe.

It's clear that the enterprise has entered into the era of big data and analytics. In addition to confirming some of our suspicions, we've uncovered a few surprises too.

## KEY FINDINGS

### 1. Multi-Cloud Strategies Are The Reality For Most Companies

Most are going with a hybrid/multi-cloud strategy with only **24% choosing to go with a single cloud vendor.**

### 2. Investment In Hadoop Is Staying The Same Or Growing

**Hadoop investment is staying the same for 53%** of respondents while 30% of respondents plan to invest more in Hadoop.

### 3. More Companies Are Implementing Data Virtualization

**55% of respondents plan to invest in data virtualization** in the near future if they are not already.

### 4. Data Governance Is A Top Challenge Across The Board

**80% of respondents said that data governance is very important** to them.

### 5. Azure Is Gaining Market Share In The Public Cloud Space

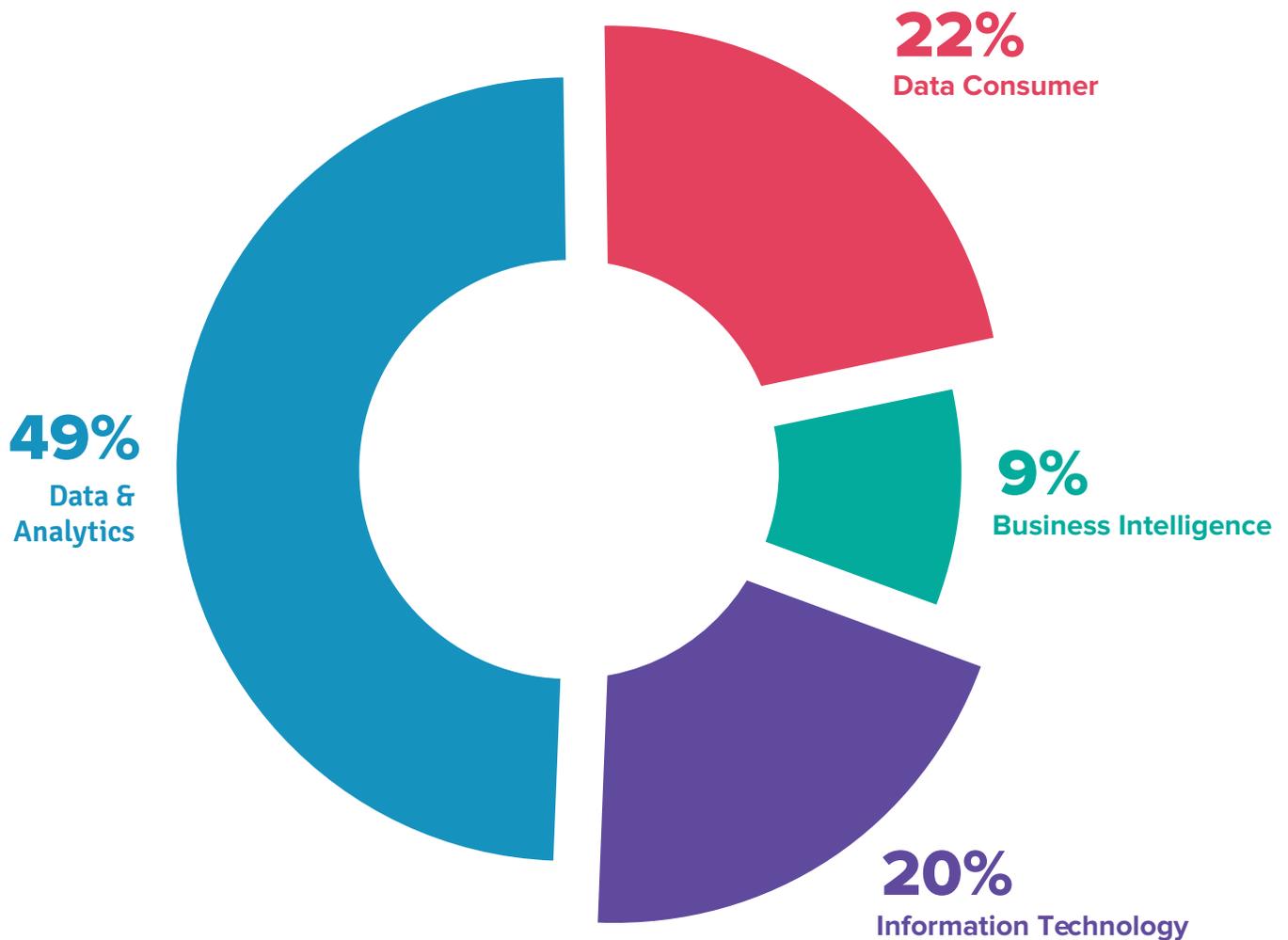
**Azure is the most popular cloud data warehouse used in our survey**, followed by AWS, Google and Snowflake.



## AUDIENCE METRICS

We collected data from over **150** data practitioners for the Big Data & Analytics Maturity 2020 Survey. The respondents were from multiple industries and based around the world. **49%** identified as data and analytics professionals while the other **51%** in the business intelligence (BI) or IT teams or identified as data consumers within business units.

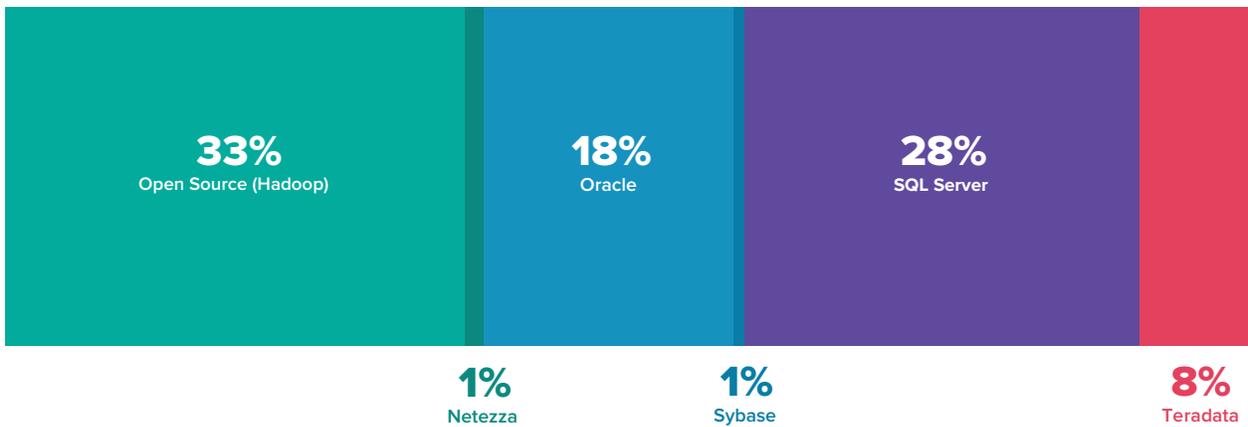
### Which department are you in?





## Which on-premise data platforms do you use?

The majority of survey respondents have data on multiple on-premise platforms. The most common sources are Open Source (Hadoop), Oracle and SQL Server. A recent CRN article asserts that “With the bulk of information technology spending still on-premises, enterprises’ modernization of their core applications will continue to transform business technology and the industry this year.” Our respondents have **33%** of their data on Open Source (Hadoop), **28%** on SQL Server, **18%** on Oracle and **8%** on Teradata.



## If you are using Open Source (Hadoop), how do you plan to invest in the future?

**53%** of survey respondents indicated that they plan to keep their investment in Open Source (Hadoop) the same. While **30%** plan to invest more.





## How do you plan to invest in your existing on-premise data platforms outside of Open Source (Hadoop)?

Respondents were more evenly split on their plans to invest in their existing on-premise data platforms other than Open Source (Hadoop) in 2020. **42%** plan to say the same. **33%** plan to invest more. And **25%** plan to invest less.



## Do you currently operate data platforms in the public cloud?

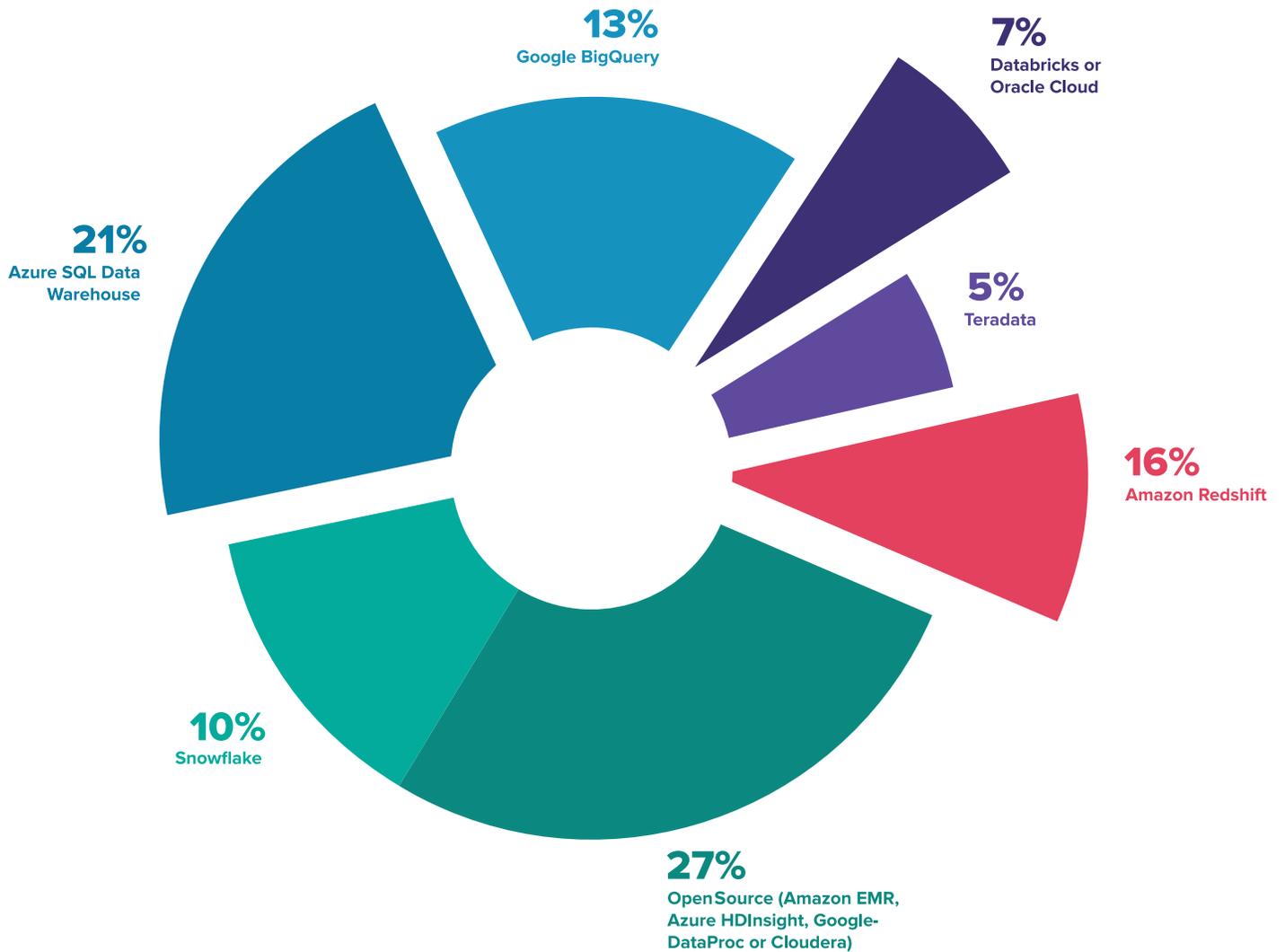
**61%** of respondents are operating data platforms in the public cloud. This is in line with [Forrester's research](#) that shows that currently, **65%** of North American enterprises rely on [public cloud platforms](#).





## Which cloud data platforms are you using?

When asked to check off which data platforms they were working with, respondents showed that they are using multiple data platforms in the public cloud from AWS to Open Source to Teradata Cloud.





## If you aren't on the cloud now, do you plan on deploying cloud data platforms in the future?

Of the **38%** of respondents who indicated that they are not on the cloud at all, **48%** plan to deploy a public cloud data warehouse/platform in the future, **16%** said that they plan to stay off the cloud, and **36%** are still evaluating their options.



## How would you rank your organization's maturity running a cloud data platform?

When asked to rank how your organization's maturity with running a data platform in the cloud, **23%** are deployed in the cloud and say that it's working well. **29%** are running a data platform in the cloud and are still working out wrinkles while **3%** say that it's not working well. The remainder of respondents said they were either not on the cloud or want to get there but haven't started planning yet.





## What is your timeline for deploying a cloud data platform?

**41%** of survey respondents are either deployed or in the process of deploying on the public cloud.

**21%** will be deployed on the public cloud in less than a year.



## What is your deployment strategy?

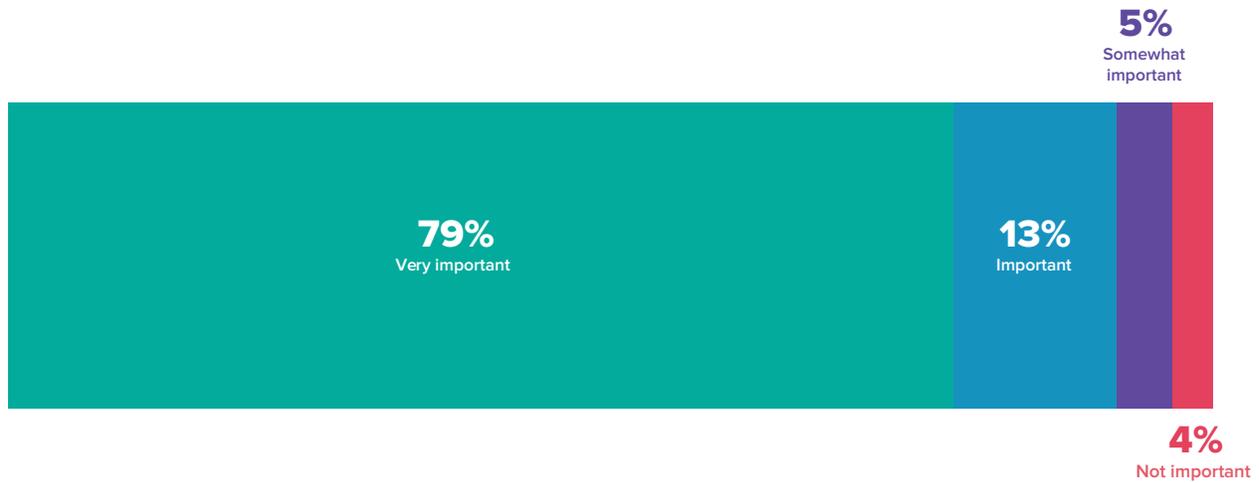
We found it surprising that only **24%** of those surveyed are all in with a single cloud vendor. The majority of respondents are working with multiple public cloud vendors and have a hybrid cloud strategy in place.





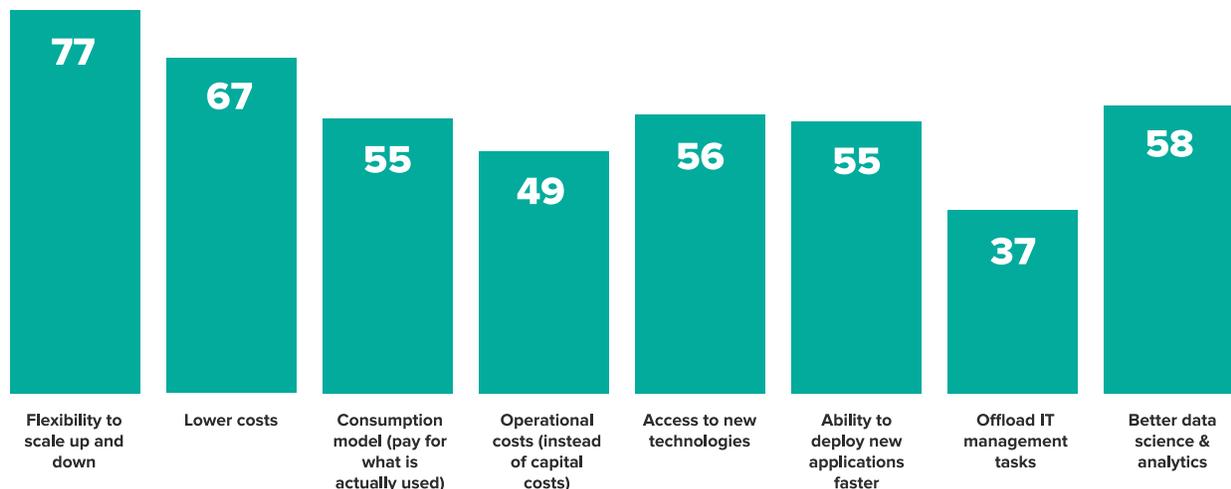
## How important is having consistent, integrated security and governance for your data in the cloud?

The vast majority of respondents **79%** indicated that having consistent, integrated security and governance for your data in public cloud, private cloud and/or hybrid cloud is very important.



## What are you hoping to achieve with the public cloud?

When respondents were asked what they were trying to achieve in the public cloud, the top responses were: **flexibility to scale up and down**, get **better data and analytics**, and **lower costs**. Close behind that was the ability to access new technologies and deploy new applications faster.





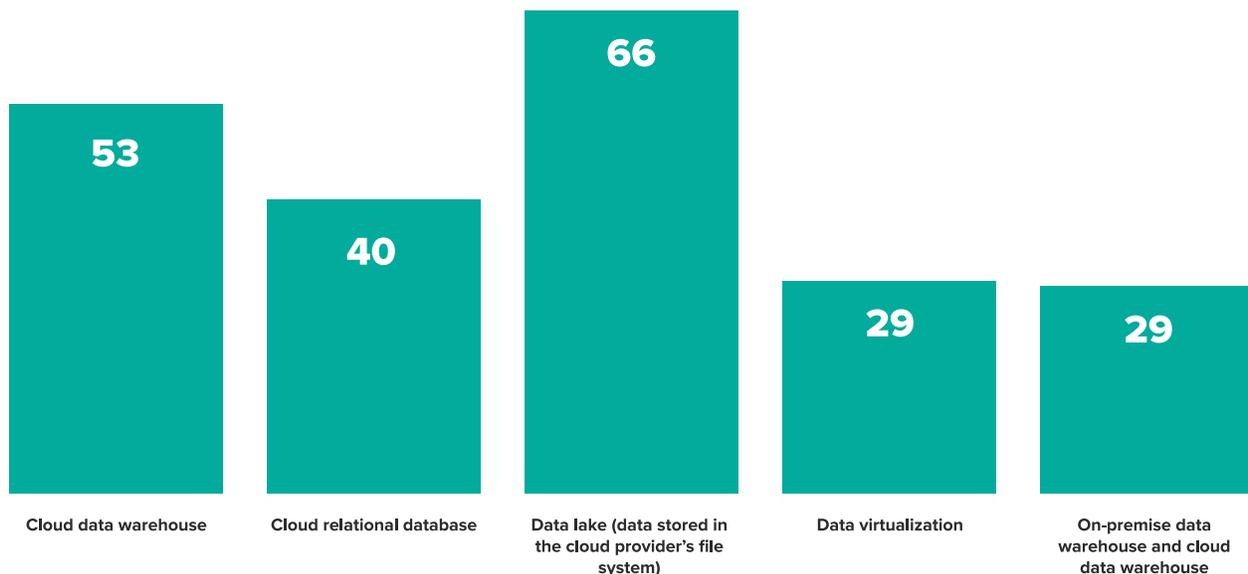
## What types of workloads are you deploying or do you plan to deploy on the public cloud?

Our respondents told us that they are looking to deploy data science, data warehousing, and business intelligence in the cloud. And how are they going to do it? By streaming real-time data and using ETL.



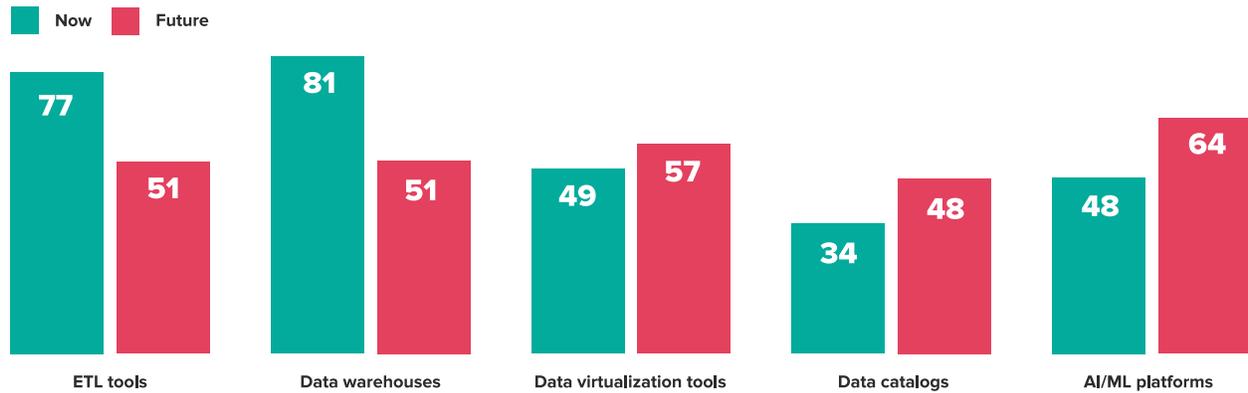
## What is your strategy for storing data in the public cloud?

The majority of our respondents are storing data on a data lake in the public cloud or in a cloud data warehouse.





## Which technologies are you using today and which technologies do you plan to use in the future?

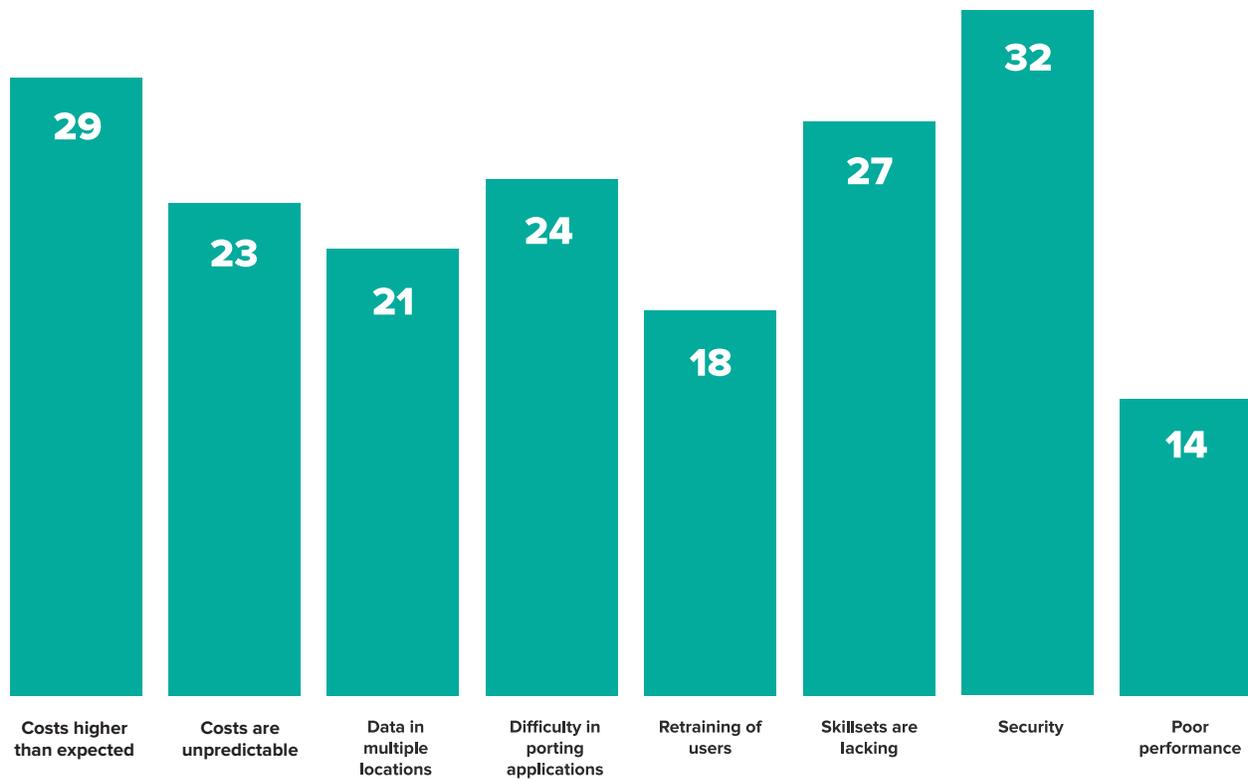


Our survey shows that **49** respondents are using data virtualization now and **57** respondents plan to use data virtualization in the future. Data virtualization technology has been around for quite some time. However, the explosion in data size and variety, coupled with the increased focus on analytical use cases, has created new challenges for legacy data virtualization technologies. The need for ad hoc access to both live and historical data has steadily increased among business users as they leverage artificial intelligence (AI) and machine learning (ML) platforms for analytics.



## What challenges are you experiencing with the public cloud?

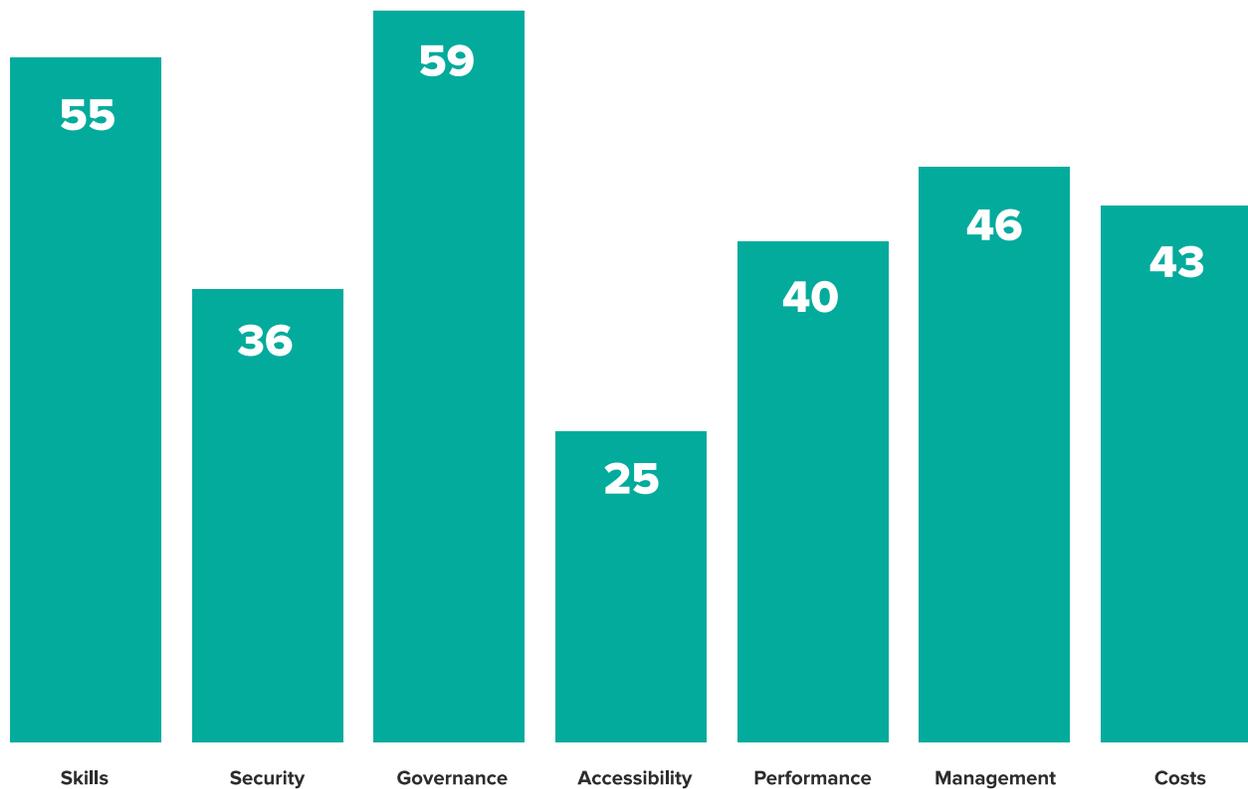
When asked about the challenges that respondents are experiencing with the public cloud, **32** people indicated that security was a challenge. Given that **79%** of respondents indicated that having consistent, integrated security and governance for your data in public cloud, private cloud and/or hybrid cloud is very important, this is not a surprise. It's also not as surprising to see that **29** people said that the cost of the public cloud is higher than expected.





## What challenges are you experiencing with your analytics infrastructure?

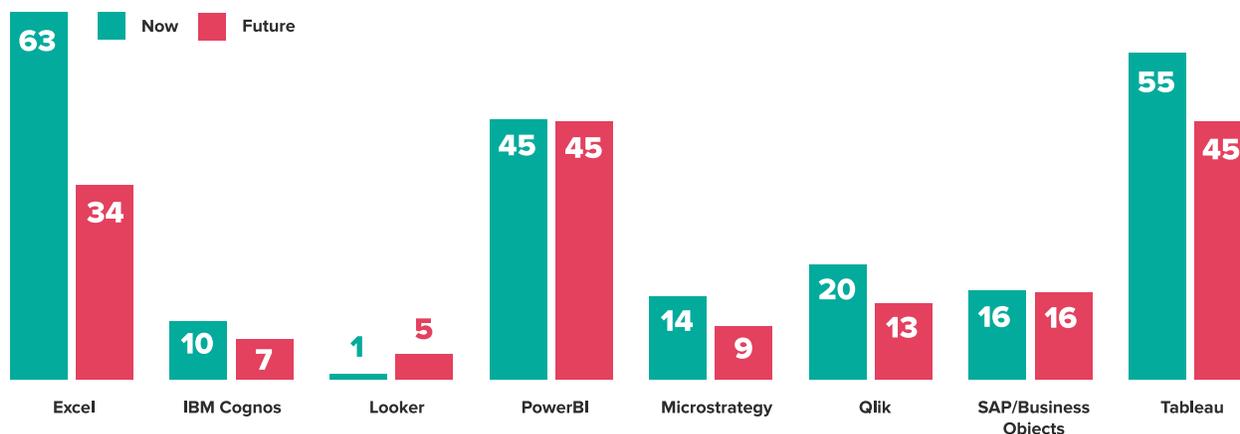
Given that **78%** of respondents indicated that having **consistent, integrated security and governance** for your data in public cloud, private cloud and/or hybrid cloud is very important, it's not a surprise that when asked to select all areas that apply to challenges that you're experiencing with your analytics infrastructure, governance was the number one challenge followed by skills, performance, costs and security.





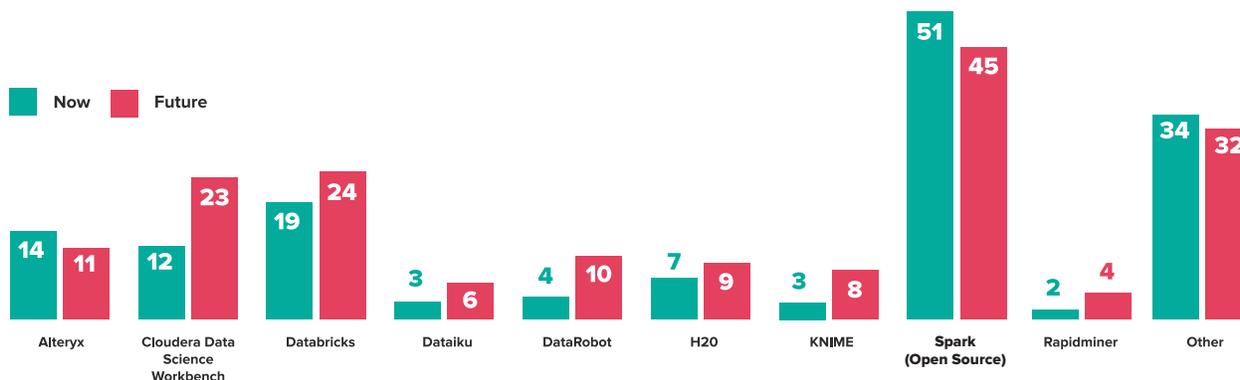
## Which BI tools is your organization using now and which BI tools do you plan to use in the future?

Microsoft applications - Excel and PowerBI - are some of the most popular business intelligence tools today along with Tableau. Looking towards the future, Excel will become less of a “go to” business intelligence tool as enterprises look to using more PowerBI and Tableau. When it comes to artificial intelligence (AI) and machine learning (ML), Spark (Open Source) is the most used technology with respondents.



## Which AI/ML tools is your organization using now and which ones do you want to use in the future?

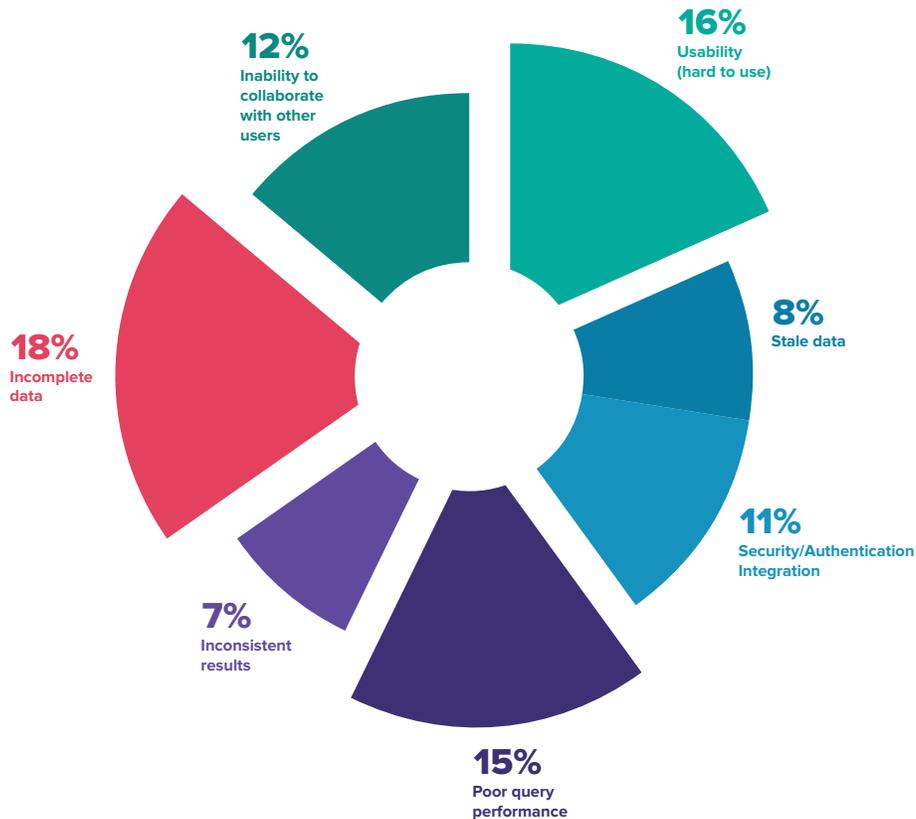
In the future, Spark remains the technology of choice, along with Databricks and Cloudera Data Science Workbench. When respondents chose “other” the most frequent responses were: Pecan, Azure, and “don’t know yet.”





## What challenges are you experiencing with your BI/AI tools?

**18%** indicated that incomplete data is the biggest challenge with their BI/AI tools while **15%** indicated that poor query performance was a big challenge.



## What would you prefer your data platform to be?

So what is the preferred data platform? **39%** said that they'd like analytics and machine learning from multiple vendors that are tied together.





## How important is having consistent, integrated security and governance across all of your analytics and machine learning functions?

**73%** indicated that having consistent, integrated security and governance across all of your analytics and machine learning functions is very important.



## CONCLUSION

Security and data governance are extremely important as enterprise customers store data on data lakes and in the public cloud and as they perform analytics and artificial intelligence on their data. And, more and more data users are looking to data virtualization in order to get live connections to data from multiple sources so that they can make better data-driven decisions for their business.