



WEKA

# The State of AI and Analytics Infrastructure 2021

Published: February 2021

# TABLE OF CONTENTS

01

HIGHLIGHTS

02

DEMOGRAPHICS

03

ARTIFICIAL INTELLIGENCE  
INITIATIVES

04

CLOUD INITIATIVES

05

USE OF CPUs AND GPUs

06

NETWORKING

07

COVID-19 IMPACT

08

ROLE & OBJECTIVES

01

# HIGHLIGHTS

# HIGHLIGHTS

## About the Survey

534 individuals across IT, data, and engineering functions participated in a survey held by WekaIO during November and December of 2020. We asked participants a wide variety of questions about their IT architectures, frameworks, databases, spending, challenges, and more.

## Key Findings

- 86% of respondents have at least one AI initiative. Most companies tend to have 2-3 initiatives, while a few even had as many as 5 initiatives.
- Use cases vary significantly by vertical, but the most cited initiatives were recommender engines, scientific visualization, and image recognition.
- The approaches to AI vary, but research and government tend to build their own models that are customized to their needs, while commercial enterprises tend to use ready-to-consume models to gain time-to-market advantage.
- In general, 50% of the data used is self-generated, but other sources contribute.
- While no respondents stated that they plan to be cloud only, most expect a significant and increasing portion of their workloads to be in the cloud. Privacy concerns and the complexity of extending workloads to the cloud are the main headwinds toward additional cloud adoption.

- While it was no surprise to see that AWS dominates the market, some industries favored other cloud providers.
- Over half of the respondents are already using GPUs, and adoption of GPUs is especially high in the Automotive segment.
- The biggest headwind to the adoption of AI initiatives is the lack of data scientists, followed closely by lack of budget and infrastructure to perform the AI.

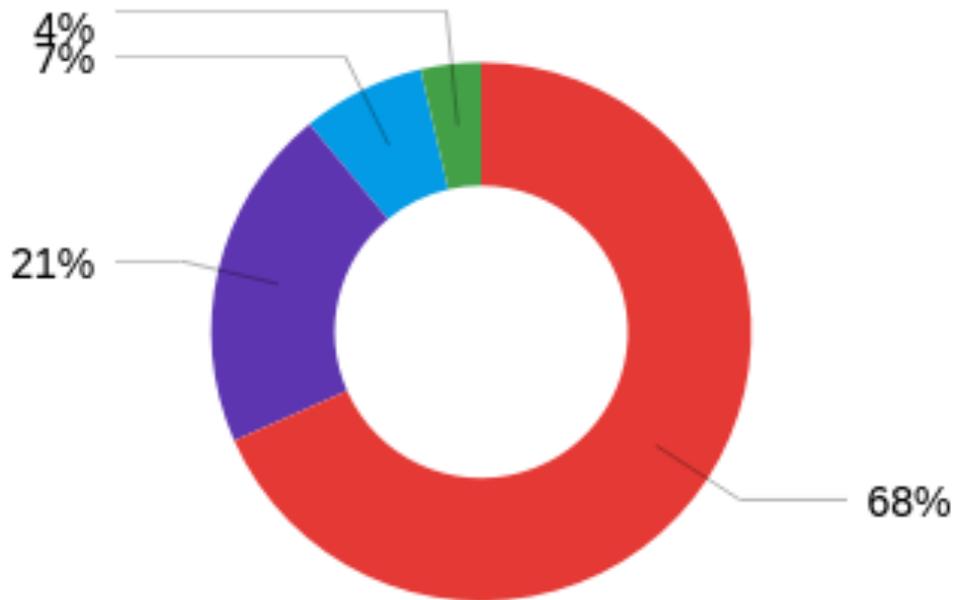
## Recommendations

- If you do not already have AI initiatives in place, you are late to the game.
- Construct a plan for processing your data well ahead of your actual need to use it. Many companies started to store data and then discovered that privacy challenges held them back from leveraging it. Other challenges when using home-grown data include the amount of cleaning and manipulation required before the data becomes useful.
- Keep your AI investment well documented. Many participants don't know how much they spend on their AI initiatives, which makes it difficult to measure the ROI or to compare whether they are spending too little or too much on their AI initiatives.

02

# DEMOGRAPHICS

# DEMOGRAPHICS BY GEOGRAPHY

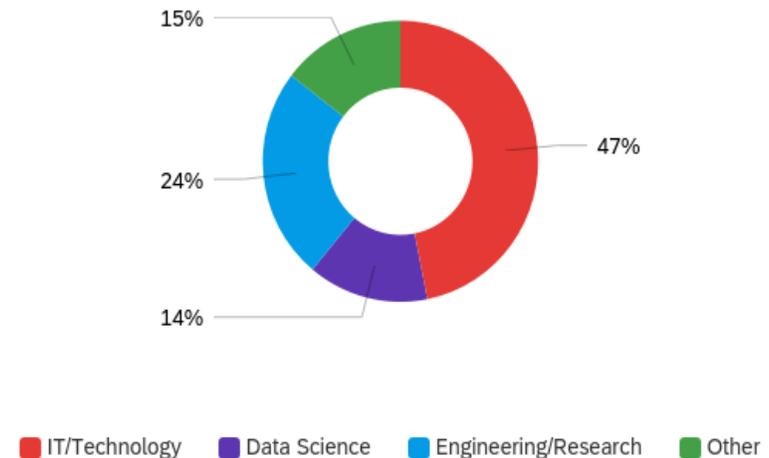
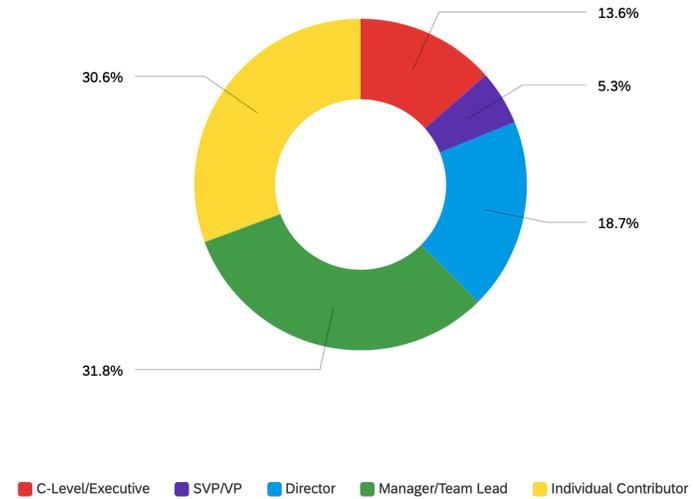


North America Europe Asia Middle East

The vast majority of participants (69%) came from North America, followed by Europe (21%), Asia Pacific (7%), and the Middle East (4%). Industry representation within the North American participants is very similar to that of the general population, with the top 3 industries in North America being Healthcare (14%), Education (11%), and Government & Research.

# DEMOGRAPHICS BY FUNCTION AND SENIORITY

Participants came from three main functions. The majority held IT/ Technology related roles (47%), followed by Engineering/ Research (24.5%) and Data Science-related positions (14%), with the rest of the participants (15%) coming from various other positions within the company. 63% of participants indicated that they were hands-on practitioners, with senior management represented by Directors (19%), C-level executives (14%), and SVP/VP (5%).



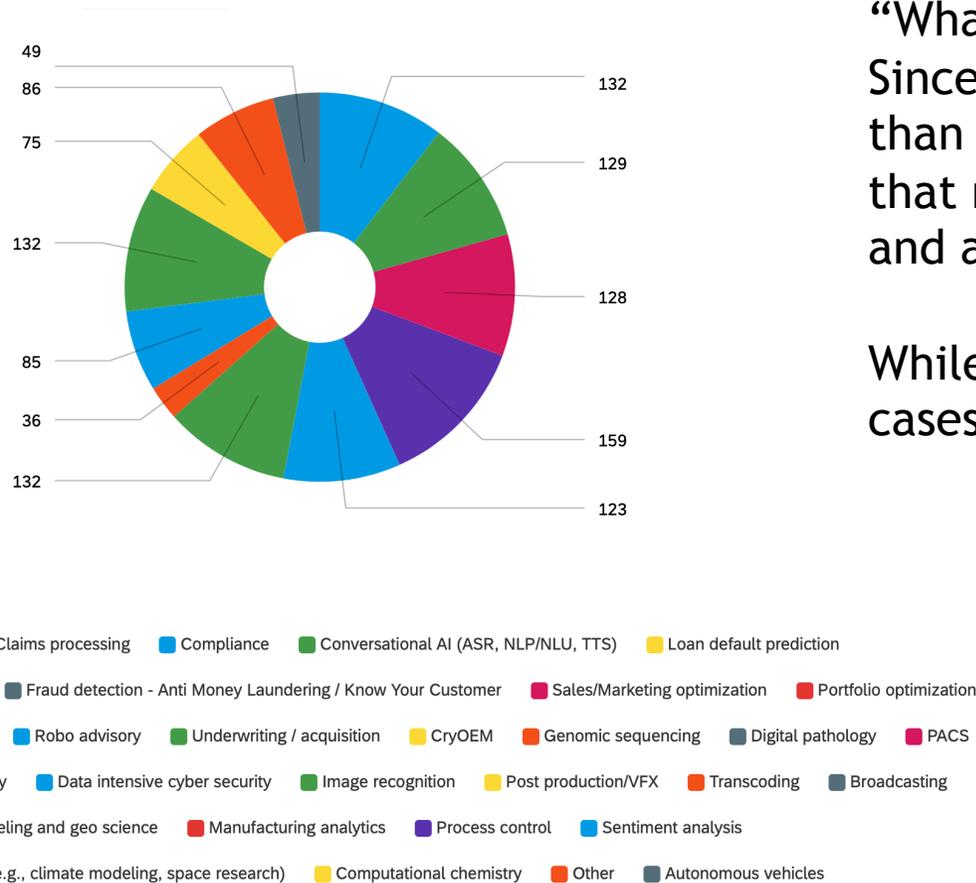
03

# ARTIFICIAL INTELLIGENCE INITIATIVES

# ARTIFICIAL INTELLIGENCE USE CASES

We received over 1000 data points on the question of “What use cases is your company investing in today?” Since it was a multiple-choice question in which more than one response was allowed, the results indicated that most companies have more than one AI initiative and are investing in multiple use cases.

While roughly 10% of the respondents have 5-6 use cases, the majority reported 2 distinct AI initiatives.



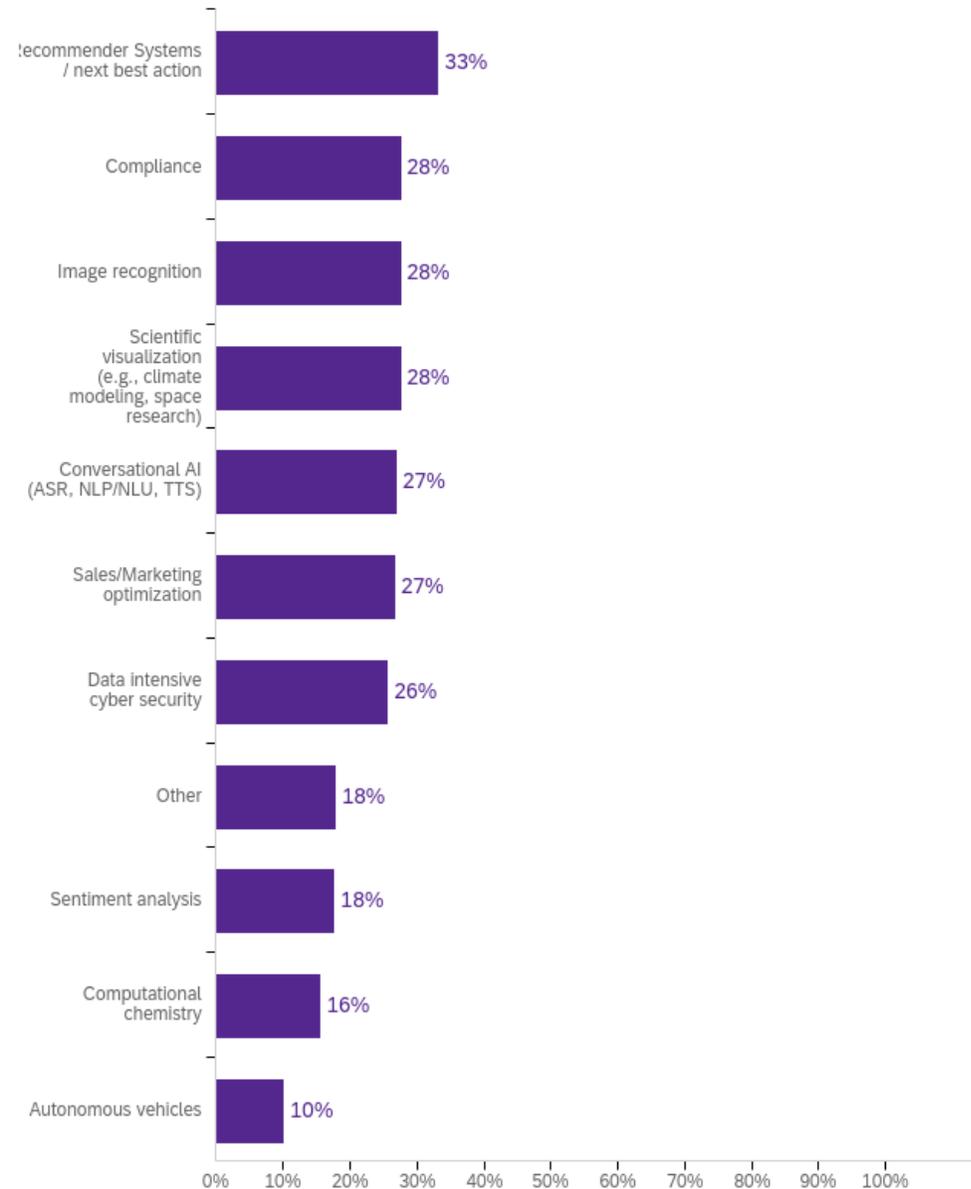
**Most companies today are investing in 2-3 initiatives**

The top use case for AI initiatives was in implementing recommendation systems. The next six are very close together, indicating similar importance of these use cases to companies. These results come as no surprise, in our view, as the early adaptors of AI within the industry seem to fit within this overall view. Over six years ago we started working with customers on such use cases as autonomous driving, image recognition, and algorithmic trading.

The survey showed that early initiatives for AI include the following:

- **Recommender systems**—Adopting recommender systems boots business, especially for high-frequency investments including instant buy or sell of financial assets based on real-time data and historical trends.
- **Scientific visualization**—Analyzing particles and computational fluid dynamics were part of this group.
- **Image recognition**—Augmenting the automotive segment, mainly autonomous transportation, with image recognition is a priority.
- **Compliance**—Proving compliance within heavily regulated environments is a high-value use case with a strong ROI that can drive quick adoption.
- **Conversational AI**—Supporting service centers to augment human interaction is a priority.
- **Transcoding**—Utilizing AI or ML can facilitate the deployment of resources more effectively but is not required for transcoding.

It will be interesting to track use cases over time to see the changes that take place. For example, we are just beginning to see the adoption of GPU-accelerated computing for computational chemistry, so we expect new use cases to grow.



# ARTIFICIAL INTELLIGENCE USE CASES BY MARKET VERTICAL

Industry	Top Use Cases	
Healthcare	Compliance	13%
	Image recognition	13%
	Scientific visualization	13%
	Computational chemistry	12%
	Cyber security	11%
Education	Scientific visualization	19%
	Image recognition	13%
	Computational chemistry	11%
	Conversational AI	11%
	Recommender systems	9%
Government and Research	Scientific visualization	24%
	Image recognition	16%
	Computational chemistry	14%
	Recommender systems	10%
	Cyber security	9%

Industry	Top Use Cases	
Cloud & MSP	Conversational AI	16%
	Recommender systems	16%
	Compliance	14%
Media & Entertainment	Sales/marketing optimization	18%
	Conversational AI	12%
	Recommender systems	10%
Telecom	Recommender systems	16%
	Compliance	13%
	Cyber security	13%
Banking & Finance	Compliance	17%
	Recommender systems	17%
	Cyber security	13%
	Conversational AI	13%
	Sentiment analysis	13%

Let's drill down to the specific use cases by industry. Depending on your specific industry, the use cases vary. For example compliance is #1 in Healthcare and Banking but not a relevant use case in Education, Research, and less regulated industries. Compliance in heavily regulated industries is a multi-million dollar opportunity given the brand impact of fraud activity that is undetected (consider the case of the Target breach in December 2013) and the potential fines by regulators. (Target was fined \$18.5M.)

Image recognition is very common in Research, given that Research tends to be dominated by High-Performance Computing (HPC) environments. In recently years we've actually seen the convergence of artificial intelligence and HPC given the vast amounts of data, mainly images, that research institutions tend to store. Research institutions use AI to run multiple iterations to test models and to gain insights. The adoption of AI in these institutions is growing significantly.

The top use cases in each vertical are those that give companies a competitive edge.

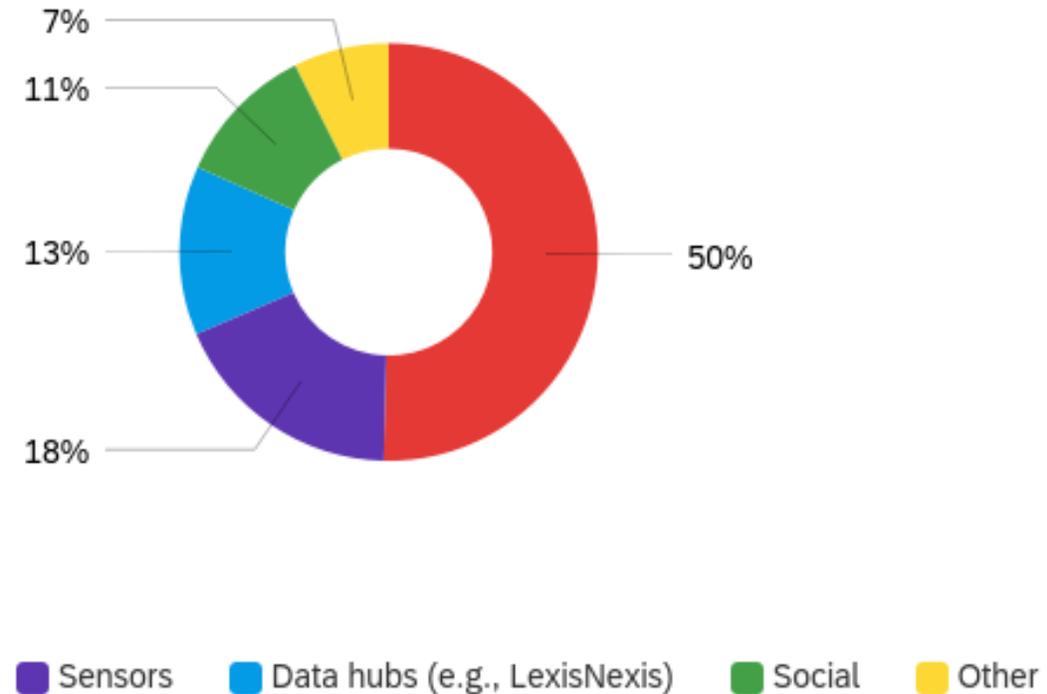
**If you are not investing in at least one of these initiatives you are losing the competitive edge that others are gaining.**

# PRIMARY SOURCES OF DATA FOR AI WORKLOADS

We were surprised to learn that 50% of the data is self generated. On the other hand, it is not surprising that 18% of the data is sourced from sensors, given the increased adoption of the Internet-of-Things (IoT) and the current ability to obtain data from even your legacy machinery.

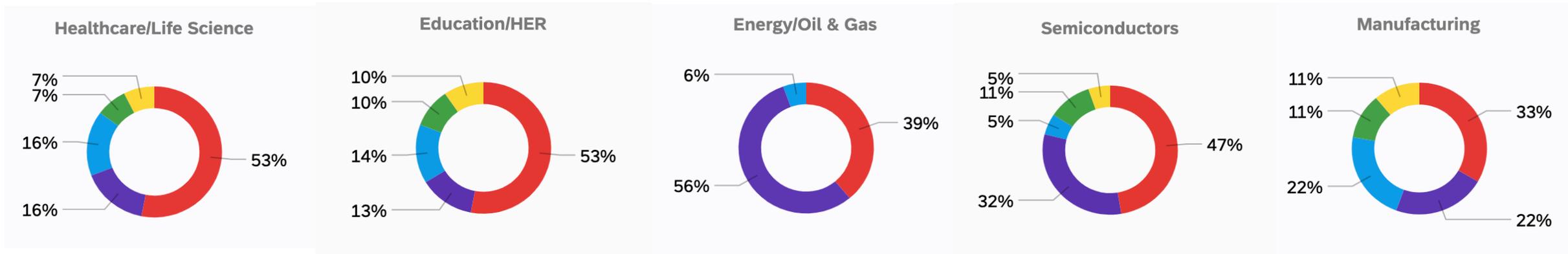
Different data sources have different requirements to manage them:

- Data coming from external data sources needs to follow an Extract, Transform, and Load (ETL) process to optimize it for analytics.
- Self-generated data requires cleaning and tagging, which should be a planned investment for the project.



As we continue to run this survey it will be interesting to see if companies begin to purchase more data and increase adoption of data hubs, as they reduce the need to prepare the data for analysis.

# PRIMARY SOURCES OF DATA FOR AI WORKLOADS BY MARKET VERTICAL



Here are some interesting findings about data sources:

- Healthcare and higher education depended significantly on self-generated data.
- The Oil and Gas and Semiconductor segments have higher amounts of data sourced from sensors and used for monitoring.
- Finance leverages external data sources, such as LexisNexis and Reuters.

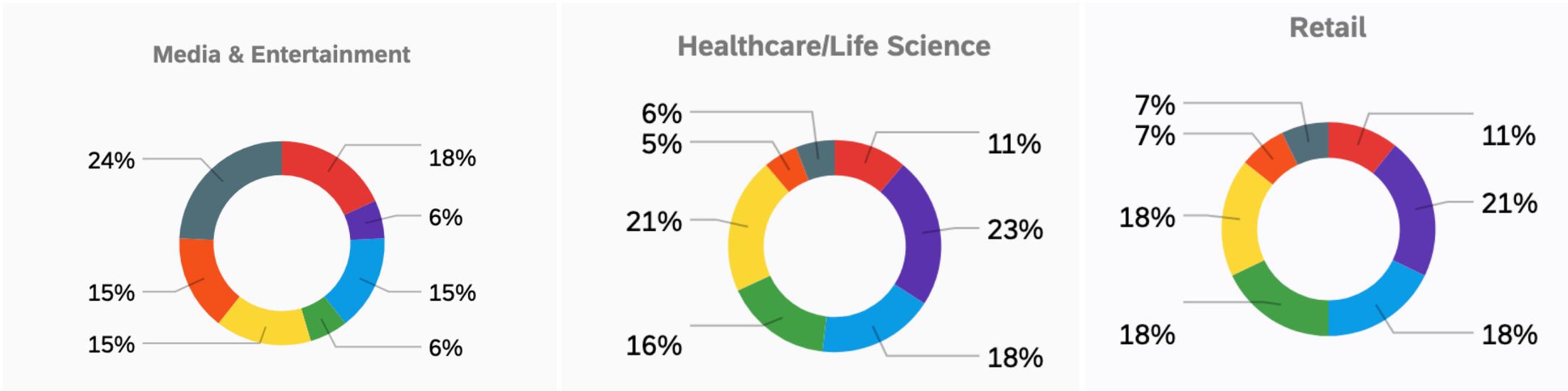
## Know in advance:

- **Where your data will come from**
- **The data's format**
- **How you will transform it**

# APPROACH TO DEVELOPING AI MODELS BY MARKET VERTICAL

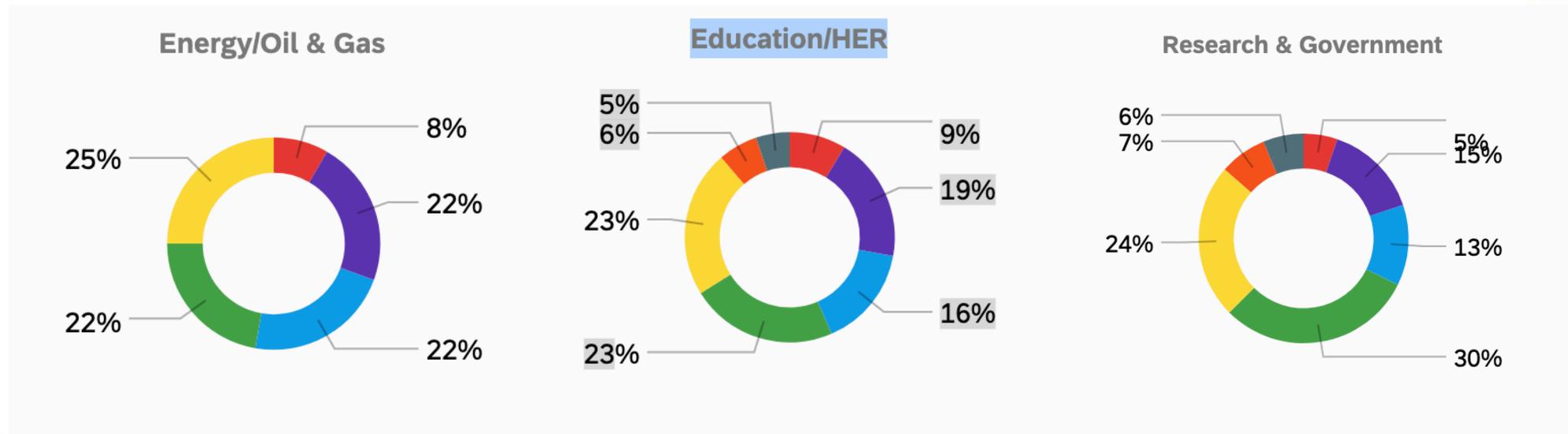
86% of responders are already adopting Artificial Intelligence and Machine Learning. The approach to developing AI models changes depending on the vertical. Overall, we can identify two approaches:

1. **Traditional enterprises** have a tendency to buy pre-built models. Their focus is on the product and deliverables.



- We have an Artificial Intelligence (AI) as a service platform
- We use high-level data science modeling tools
- We use automatic machine learning tools
- We work directly in AI frameworks like TensorFlow or PyTorch
- We work directly in Python using Pandas or Scikit-learn type libraries
- Not applicable/None of the above
- We are planning to deploy AI in the next 6-12 months

2. **Research Institutions and Education** focus on building their own solutions and are less likely to buy pre-built models. They invest more data science resources into solving the problems. The outcome is that the models are more relevant to the specific needs of the organization. Oil and gas is a highly specialized market and tends to reflect more of the HPC environment than enterprise.



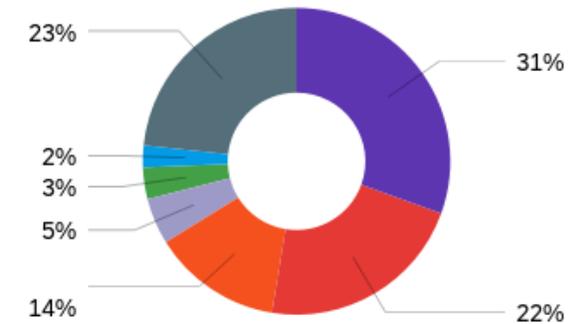
- We have an Artificial Intelligence (AI) as a service platform
- We use high-level data science modeling tools
- We use automatic machine learning tools
- We work directly in AI frameworks like TensorFlow or PyTorch
- We work directly in Python using Pandas or Scikit-learn type libraries
- Not applicable/None of the above
- We are planning to deploy AI in the next 6-12 months

# MOST COMMON FRAMEWORKS USED FOR AI MODELS

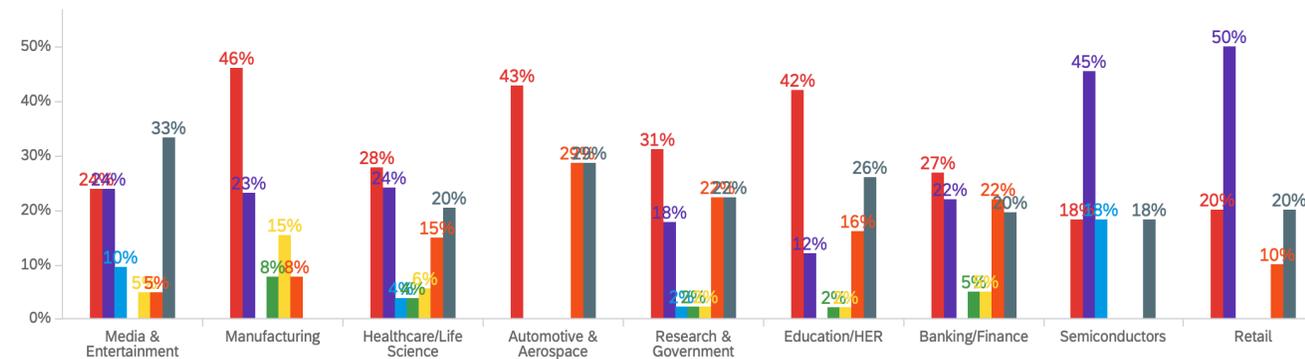
There are multiple ways to start your AI initiative : from self-built models vs. buying existing models. When developing your models there are multiple choices. Some of the more common ones:

- TensorFlow & PyTorch - Lower level frameworks used to process large data. Tend to offer better performance than Keras.
- Keras - easy to use but tends to be low performant. A great way to start. Can be used on top of TensorFlow.
- Café - Used mainly in the edge. Edge AI is just starting to evolve. Instead of pushing data to the core for analysis, analysis is done at the edge. We anticipate the adoption of Edge AI to grow along with the adoption of edge computing.

An analysis by verticals demonstrates that PyTorch is more common in semiconductor and Retail, while TensorFlow is more common in research, banking and healthcare.

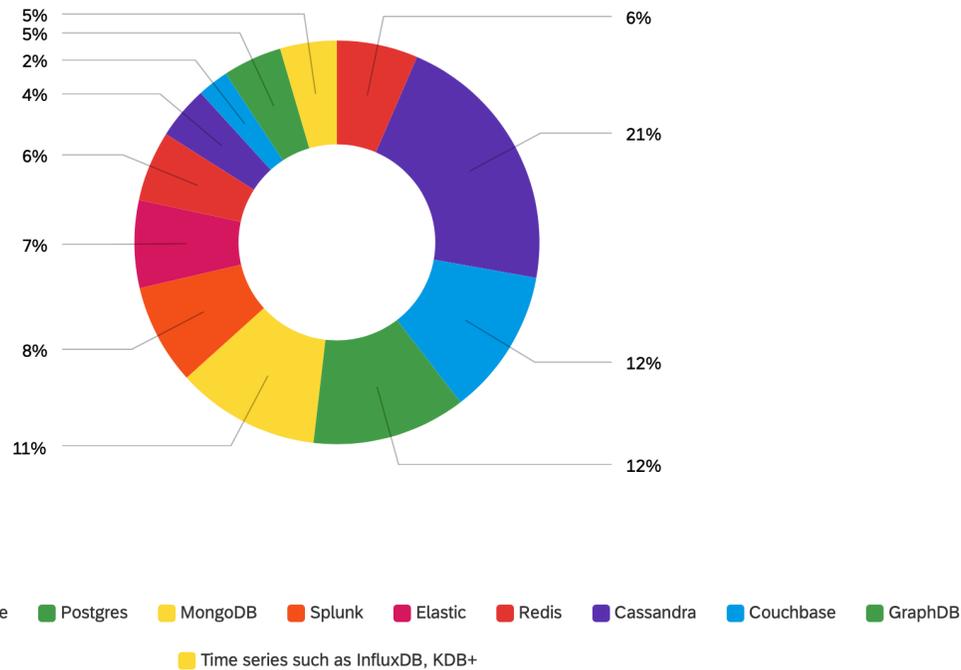


Legend for Donut Chart: TensorFlow (Purple), PyTorch (Red), NumPy (Orange), Keras (Light Purple), Caffe (Green), MXNet (Blue), Other (Grey)



Legend for Bar Chart: TensorFlow (Red), PyTorch (Purple), MXNet (Blue), Caffe (Green), Keras (Yellow), NumPy (Orange), Other (Grey)

## MOST COMMON DATABASES IS YOUR COMPANY USING

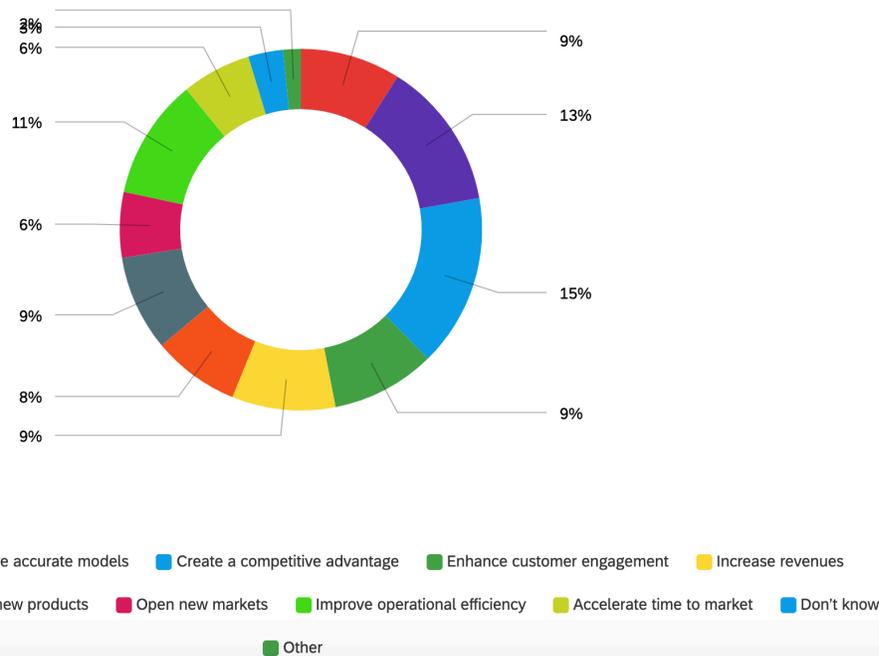


Rational databases such as MySQL and Oracle account for about a 3<sup>rd</sup> of the responses. Postgres and MongoDB gaining momentum.

It is no surprise that rational databases such as MySQL and Oracle are more common in industries that have been gathering data for many years such as healthcare and media, while industries that are leveling IOT data, such as semiconductor and oil and gas tend to use nonrational databases such as MongoDB.

We were surprised to see more balance between the various databases in banking and attribute that to both the new use cases and how competitive the market is which is driving innovation.

# WHAT IS THE MEAN DRIVER BEHIND YOUR AI INITIATIVE



The #1 reason for investing in AI is to gain a competitive advantage. This is consistent with other research conducted by the Altimeter group that revealed that 93% of companies experiencing high growth had invested in AI initiatives, while 64% of single digit growth companies had no AI initiative in place.

9% attributed the driving to creating new products which can be viewed as another form of creative competitive edge. Together they account for a quarter of the respondents.

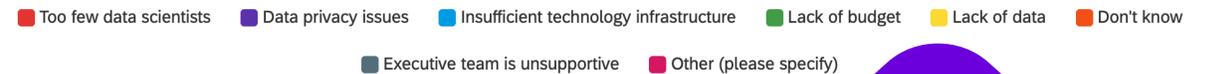
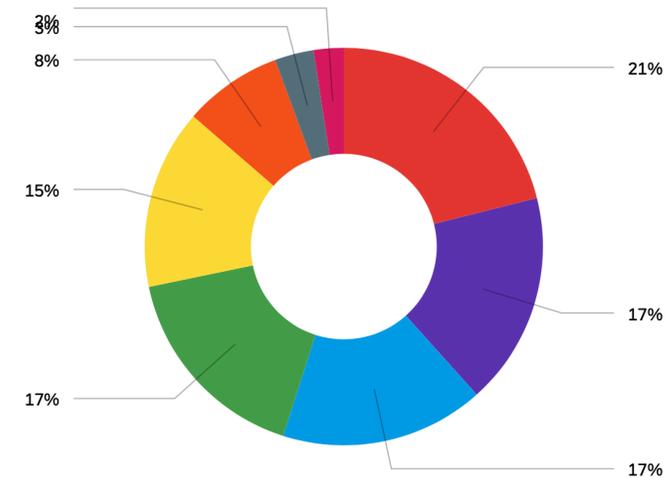
This is consistent with what we see within our customer base. Some of the companies we've been working with for a few years now and leapfrogging the market.

## WHAT ARE THE BIGGEST HEADWIND IN ACHIEVING YOUR AI GOALS

We were somewhat surprised to learn that not enough data scientists was the biggest headwind in making AI initiatives successful.

As a company in the middle of AI initiative you are also competing on human capital. Make sure your environment is inviting.

The challenges around data privacy came as no surprise. Different geographies have different privacy policies. That's becoming a critical point. When you start to collect data you want to make sure you think about how and where you plan to use it and make sure you are compliant. Think about the data that you are collecting on day 1 to avoid privacy challenges down the road.



**BEFORE  
COLLECTING  
DATA THINK  
THROUGH HOW  
YOU ARE  
PLANNING TO  
PROCESS IT.**

04

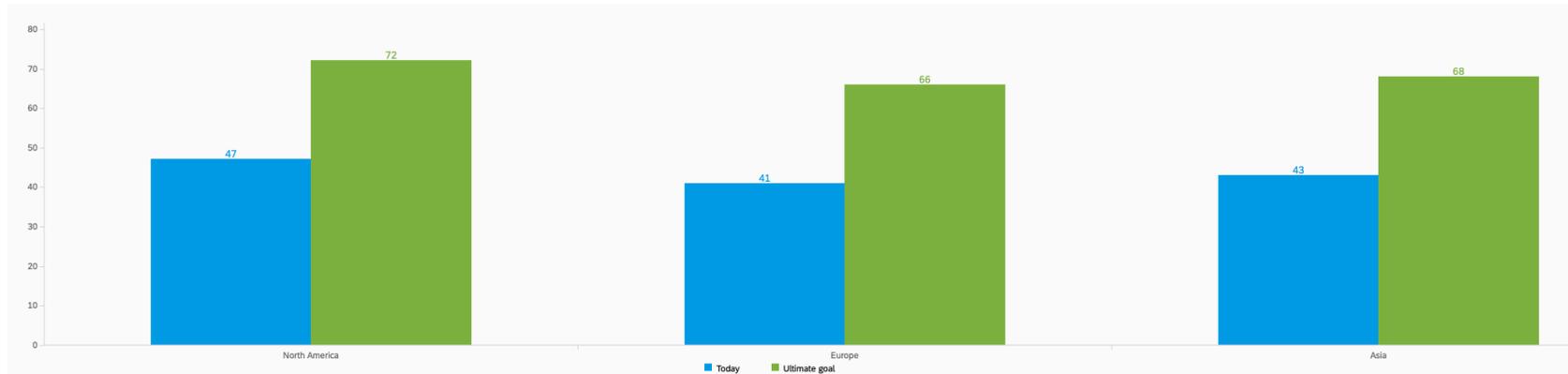
# CLOUD INITIATIVES

# PERCENTAGE OF WORKLOADS IN THE CLOUD

Across the board, respondents indicated heavy usage of the cloud today (~46%) and a consistent global path for ultimately transitioning more workloads to the cloud.

Ultimately, the cloud will play an even larger part in infrastructures moving forward (~70%), whether they be in AI or general use. Nevertheless, companies show the goal of retaining ~30% of their core workloads on premises.

Now let's see the challenges that organizations are facing in meeting their cloud objectives. There was little regional difference, indicating that cloud adoption is a global phenomenon.

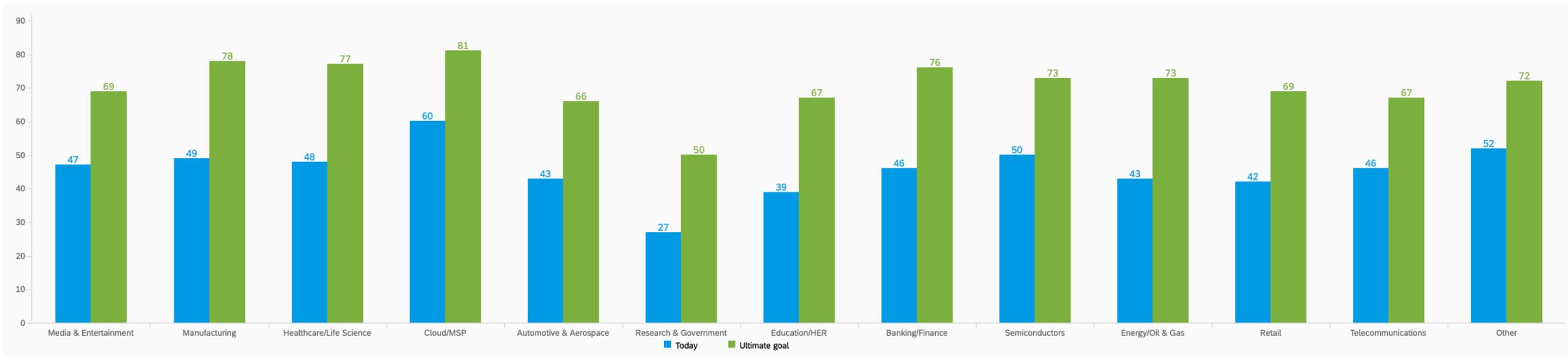


**Enterprises are unlikely to be All Cloud**

# CLOUD ADOPTION BY VERTICAL MARKET SEGMENT

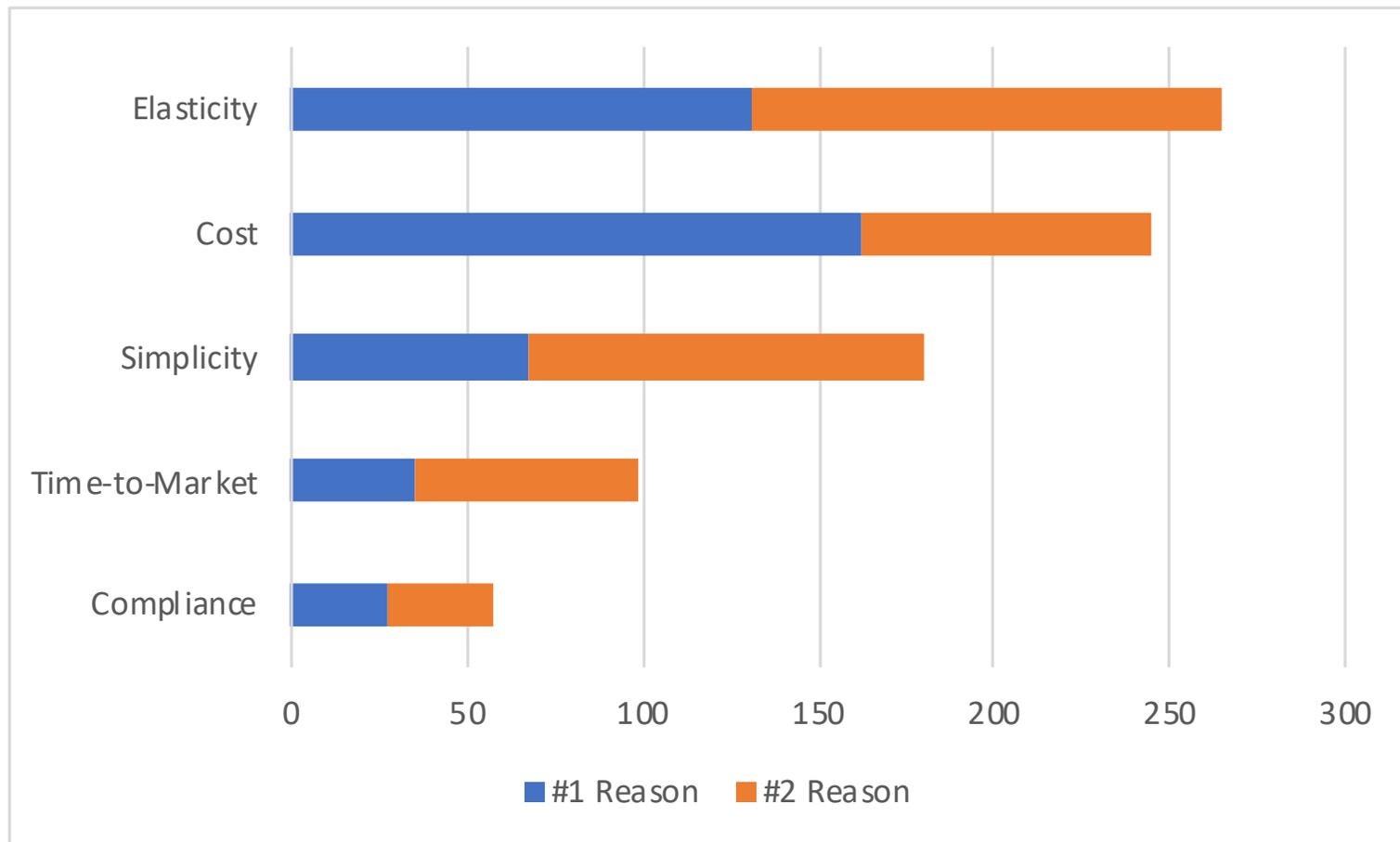
We expected higher cloud adoption rates from Media & Entertainment, which have graphic intensive workloads and are well suited to the bursty nature of the cloud. Research and government have a significantly lower cloud adoption rate than other verticals, which is not surprising given their stringent regulations and concern for data security.

One remark to note: Not all clouds are the same. There is a rise in traditional vendors offering specialized cloud environments that are different from enterprise public clouds, such as AWS GovCloud. There are also clouds like Seismic cloud for data sets in Oil & Gas.



# TOP REASONS FOR MOVING TO THE CLOUD

Respondents indicated that cost was their #1 Reason for moving to the cloud. However, research validates that when we combined the #1 and #2 reasons, customers view cloud elasticity as an even bigger influence overall. It comes as no surprise. The ability to scale from 100 CPU cores to 1000 cores in minutes and scale back once done is unparalleled.

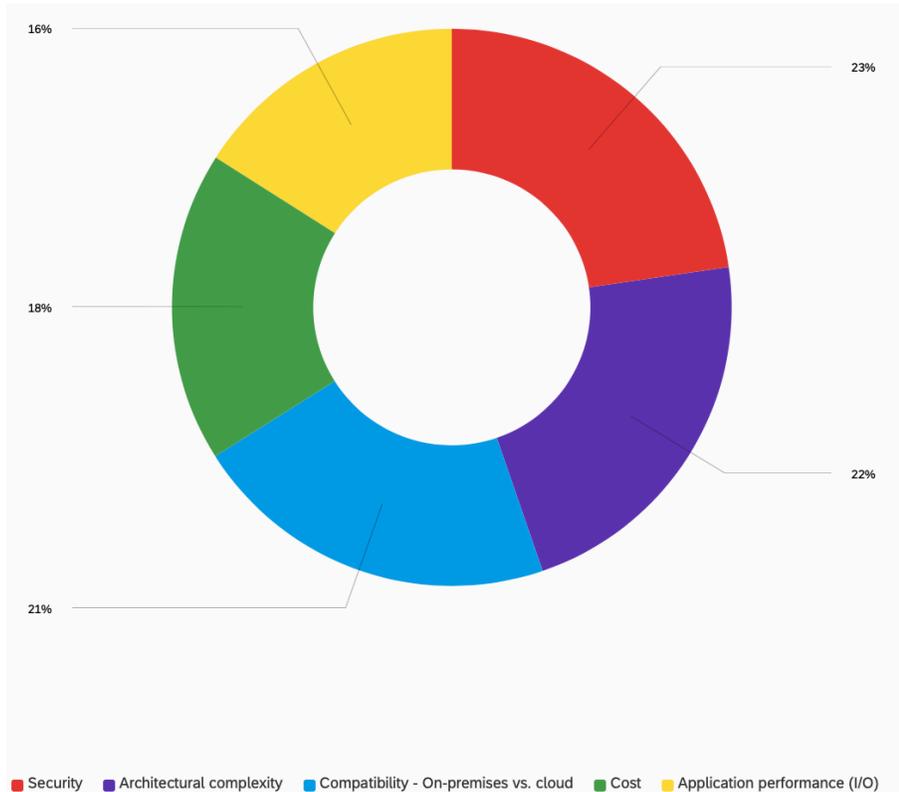


We observe that elasticity and cost are more easily quantified and measured, so they are more well known.

We actually consider elasticity as one sort of cost consideration since the cloud allows bursting without the need for huge CAPEX investments required for on-premises.

The bottom line is that the ability to maintain agility and grow is key for customers as their datasets and workloads continue to increase in size.

# HEADWINDS TO MIGRATING WORKLOADS TO THE CLOUD



Security is the #1 reason for holding back movement to the cloud, especially for manufacturing given their IP.

Complexity remains a concern, as companies wonder how to begin extending workloads to the cloud.

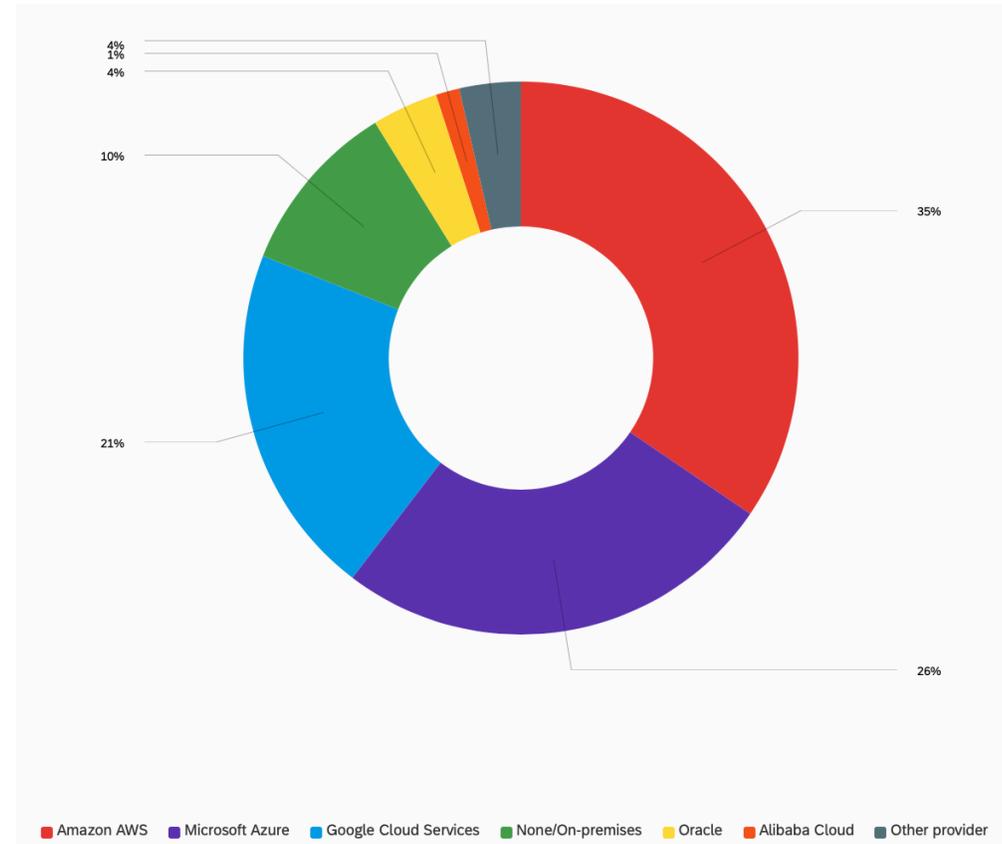
In our view, given that there are so many AI-as-a-Service offerings in the cloud, the starting point is the main barrier, and companies that begin in the cloud achieve much faster results.

Cloud performance is no longer a large issue, given the introduction of GPUs, next-gen storage solutions, and newer networking technologies.

**Privacy and architectural complexity are the main headwinds to cloud adoption**

# PRIMARY CLOUD VENDOR OF CHOICE

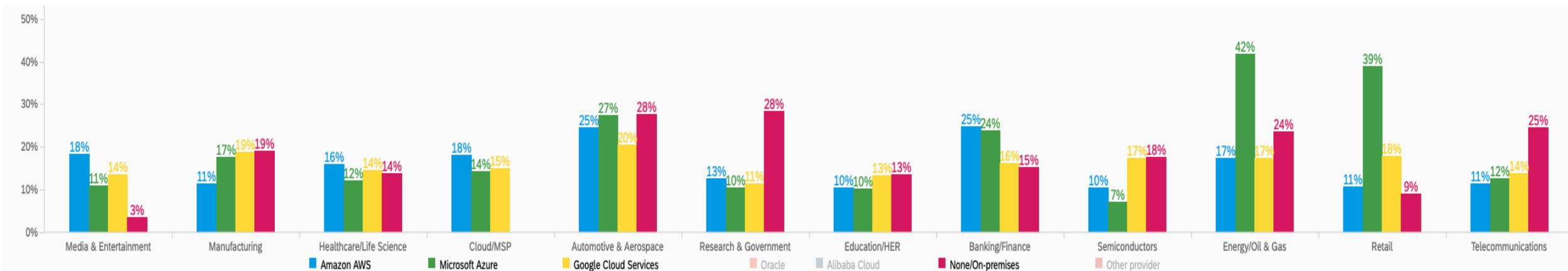
Survey results were consistent with the market overall, so there were no surprises here with AWS gaining 35% market share, followed by Microsoft Azure with 26% and Google Cloud Platform with 21%. The results tend to offer more insight when analyzed by vertical.



# PRIMARY CLOUD VENDOR OF CHOICE BY VERTICAL

It's not surprising that Amazon AWS is dominant in Media & Entertainment, given the popularity of VFX and Studio in the Cloud. Azure seems more popular in Banking, Energy, and Retails, probably due to the long-term relationships that Microsoft has had with those industries.

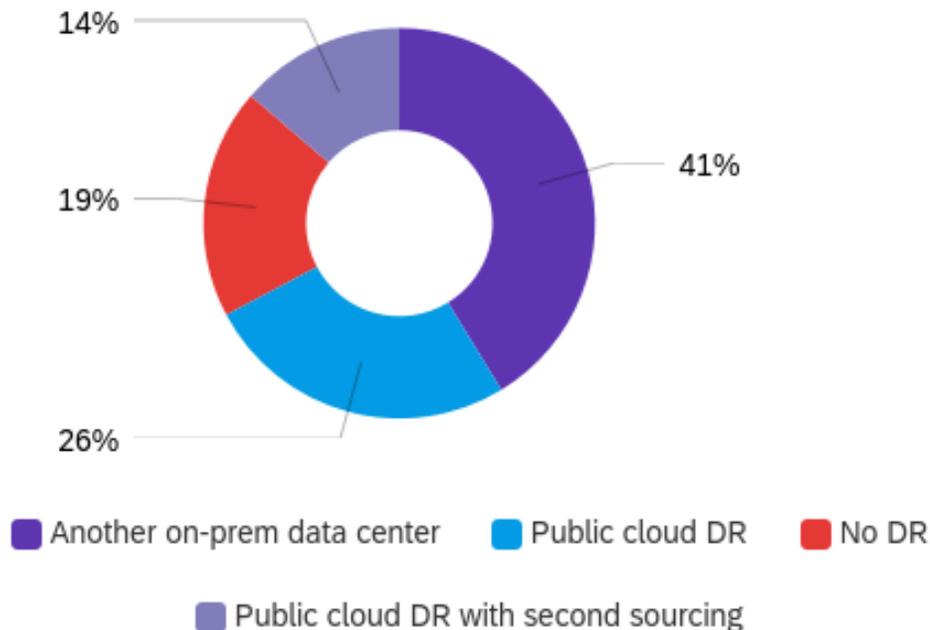
Research and Government are more inclined to use predominantly on-premises solutions given the sheer size of their data sets and the sensitivity of their data.



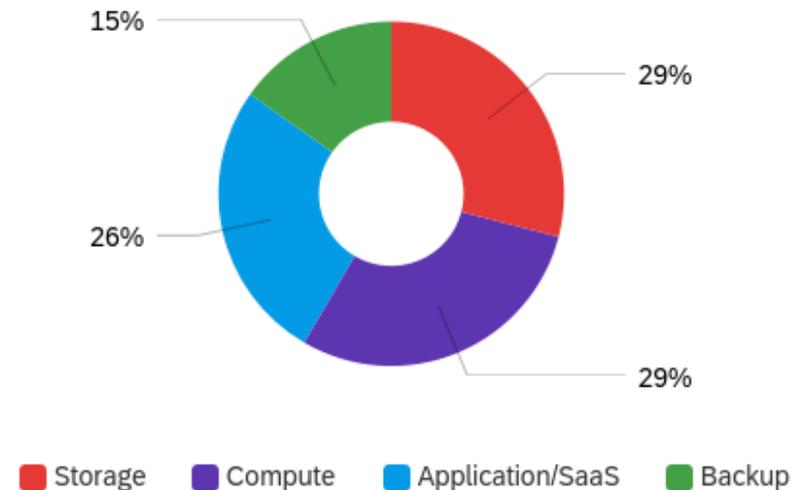
# WHAT IS THE MAIN USAGE OF THE CLOUD?

The results split evenly between compute (29%), storage of data (29%) and cloud as Software as a Service (SaaS) (26%). We expected backup to be more dominant, but only 15% of the respondents mentioning using the cloud as a backup. These results correlate with the results to the question regarding the Disaster Recovery (DR) strategy: Over 40% of respondents have an on-prem DR site, with 19% stating they don't have DR, which excludes 60% of the respondents from using the cloud for backup.

Main use case for the cloud



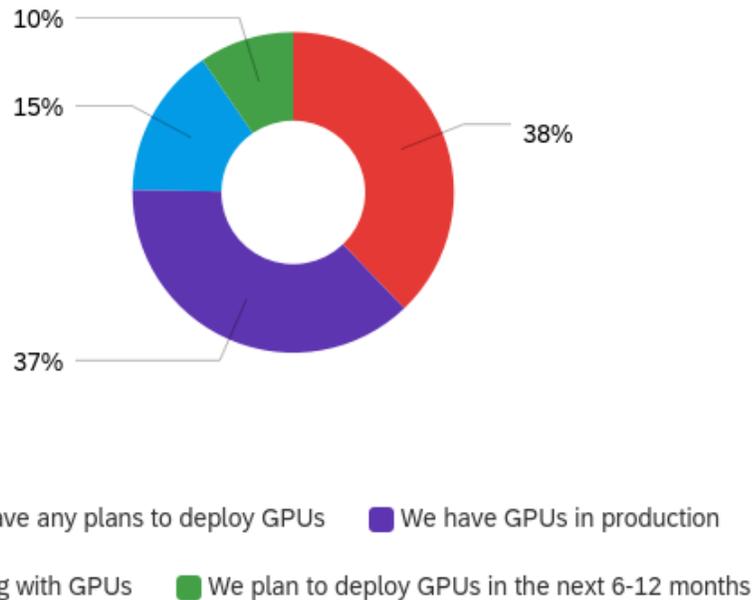
Disaster Recovery Strategy



05

# USE OF CPUs AND GPUS

# GPU ADOPTION IN AI WORKLOADS



Over half of the respondents (52%) mention using GPUs in production or pilot programs. 38% mentioned that they do not have immediate plans to deploy GPUs. Most of these respondents came from the Middle East.

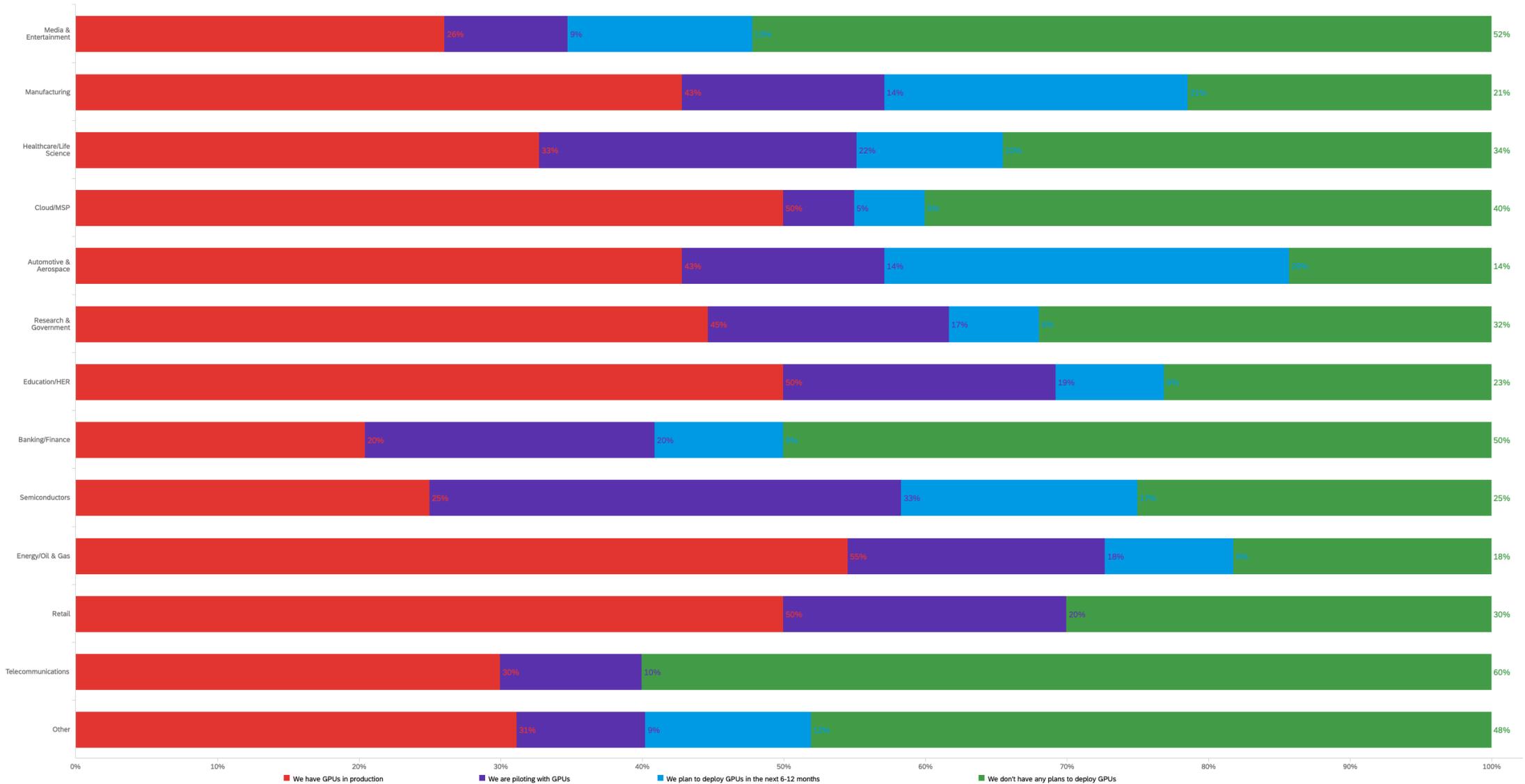
Adoption of GPUs is especially high in Automotive, where image recognition is used to power autonomous transportation. Only 14% of Automotive respondents do not have plans to use GPUs. Other industries where the addition of GPUs is above 50% include Oil and Gas/Energy, Retail, and Cloud/MSP, mainly for reselling.

Within the next 6-12 months, the verticals most likely to grow their usage of GPUs are Automotive and Manufacturing, most likely for preventive maintenance.

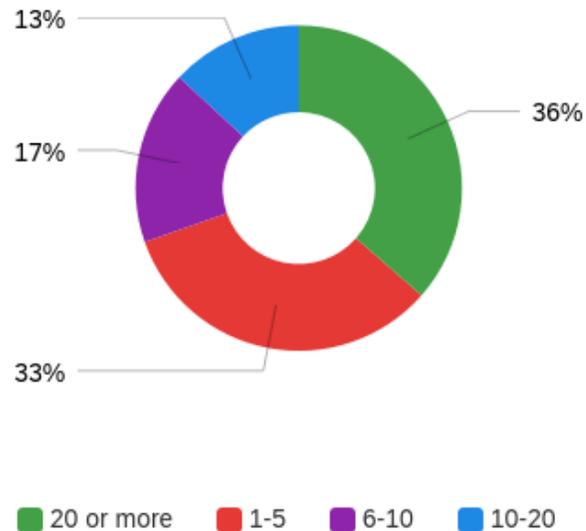
Respondents in Telecom indicate that they are the least likely to adopt GPUs, with over 60% responding against growth in that area.

**Over half of the respondents mention using GPUs.**

# BREAKDOWN OF GPU ADOPTION IN AI WORKLOADS



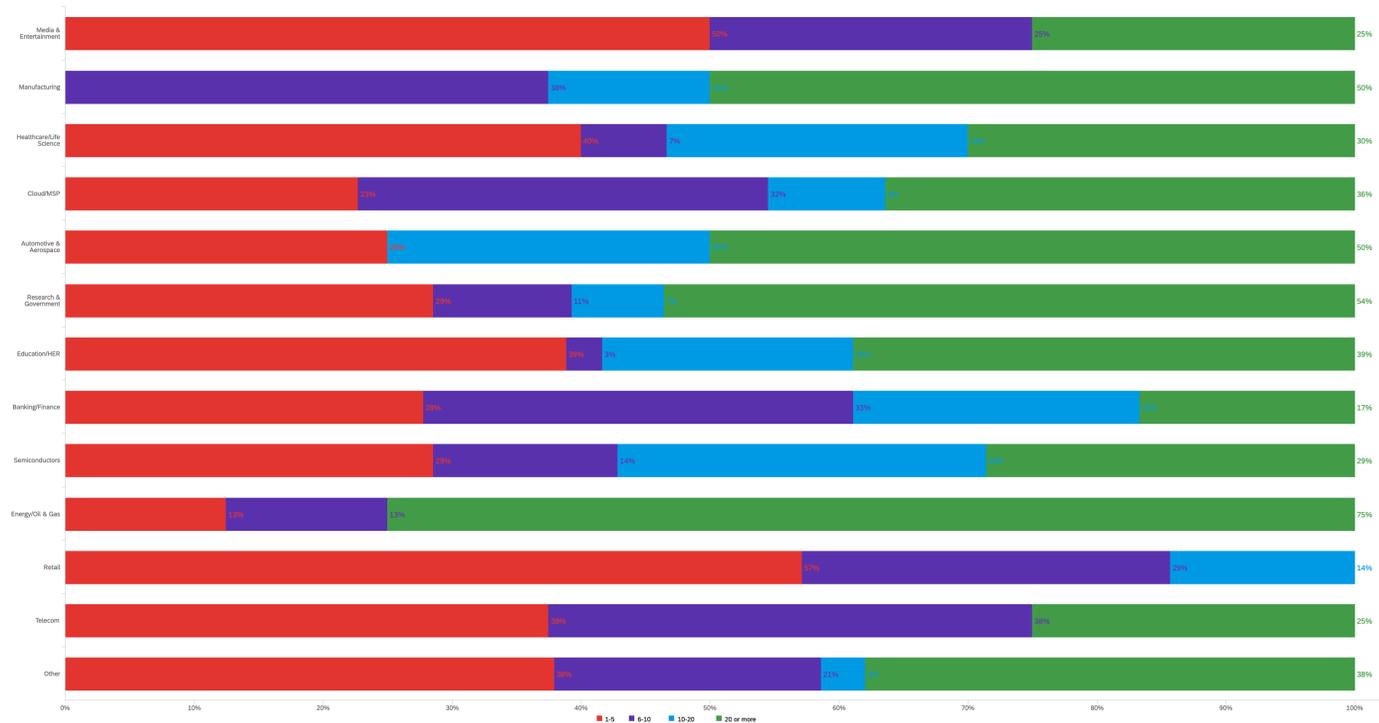
# NUMBER OF GPU SERVERS OR ACCELERATORS



Retail, Banking, and MSP, on the other hand, present numbers that are significantly lower than other industries.

It was encouraging to see that half of the industry have more than 10 GPUs/accelerators, with over a third of the respondents using 20 or more GPUs.

Adoption by industry shows correlation with the previous finding: Manufacturing, Oil and Gas, and Research presenting larger volumes of GPU servers and accelerators.



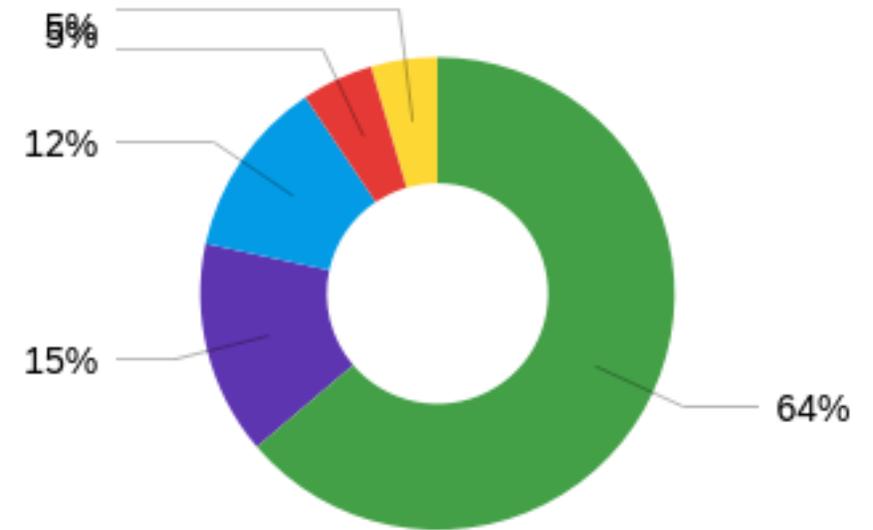
# GPU ADOPTION BY VENDOR

It comes as no surprise that NVIDIA dominates the market space. Given their history in graphics processing and the predominance of visualization/image recognition as early adopters of AI, NVIDIA is a natural fit.

Indeed NVIDIA shows dominance in Research(71%), Education (65%), and Oil and Gas (89%). All three are very familiar with HPC infrastructures and technologies, where NVIDIA first focused.

AMD is now in a strong #2 position with their recent acquisition of Xilinx, with 20% market share.

9% of the market are still using FPGAs.

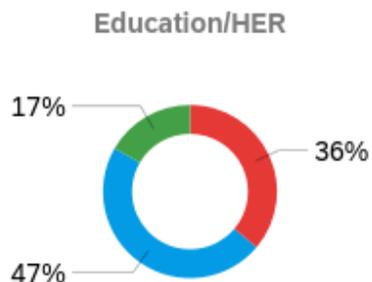
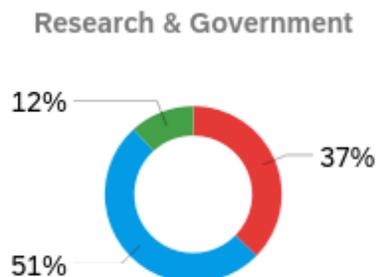
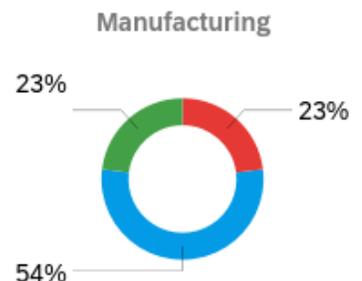


06

# NETWORKING

# WHAT PREDOMINANT NETWORK INFRASTRUCTURE CONNECTS YOUR CPUs/GPUS?

## Mostly InfiniBand and OmniPath



Overall, 54% use Ethernet, 23% use InfiniBand, and 6% use OmniPath.

In more enterprise environments, such as Banking, Media, and Healthcare, Ethernet is dominating the market.

Research institutions, Manufacturing, and Education are more likely to use InfiniBand or OmniPath.

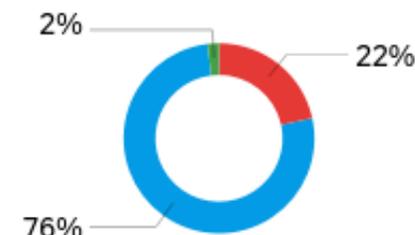
Historically, we've seen InfiniBand and OmniPath to be correlated with High Performance Computing (HPC) environments. Education and Research were among the first to adopt HPC, so this comes as no surprise.

## Mostly Ethernet

### Media & Entertainment



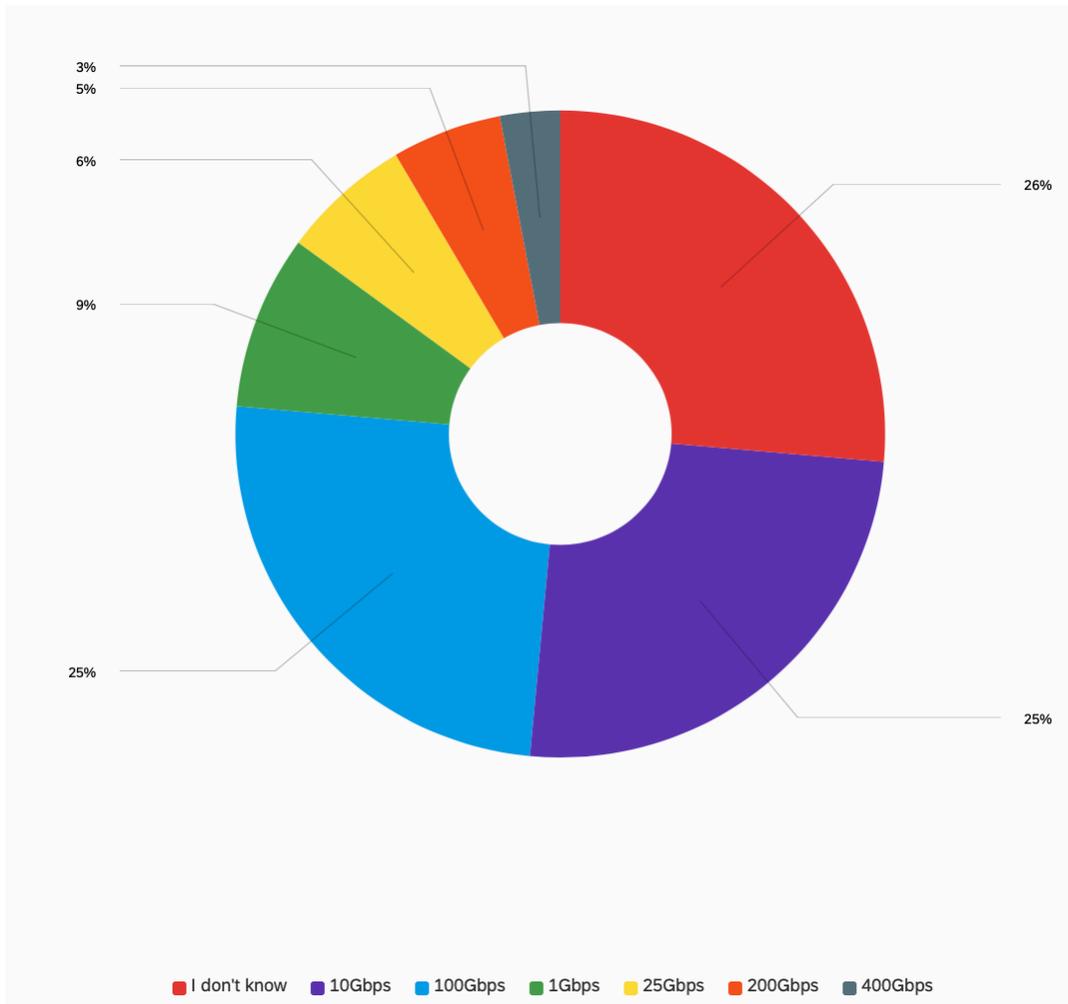
### Healthcare/Life Science



### Banking/Finance



# WHAT PREDOMINANT NETWORK SPEED CONNECTS YOUR CPUs/GPUS?



100Gbps is the predominant technology. Close to half (45%) of respondents mentioned using networks that are 100Gbps networks or faster.

With the introduction of GPUs and later NVIDIA® GPUDirect®, which were historically running on InfiniBand, the fact that they are not supporting Ethernet significantly boosts their adoption. We see that by combining the data about the networking speed, networking infrastructure, and the adoption of GPUs.

25Gbps was a transitional technology.

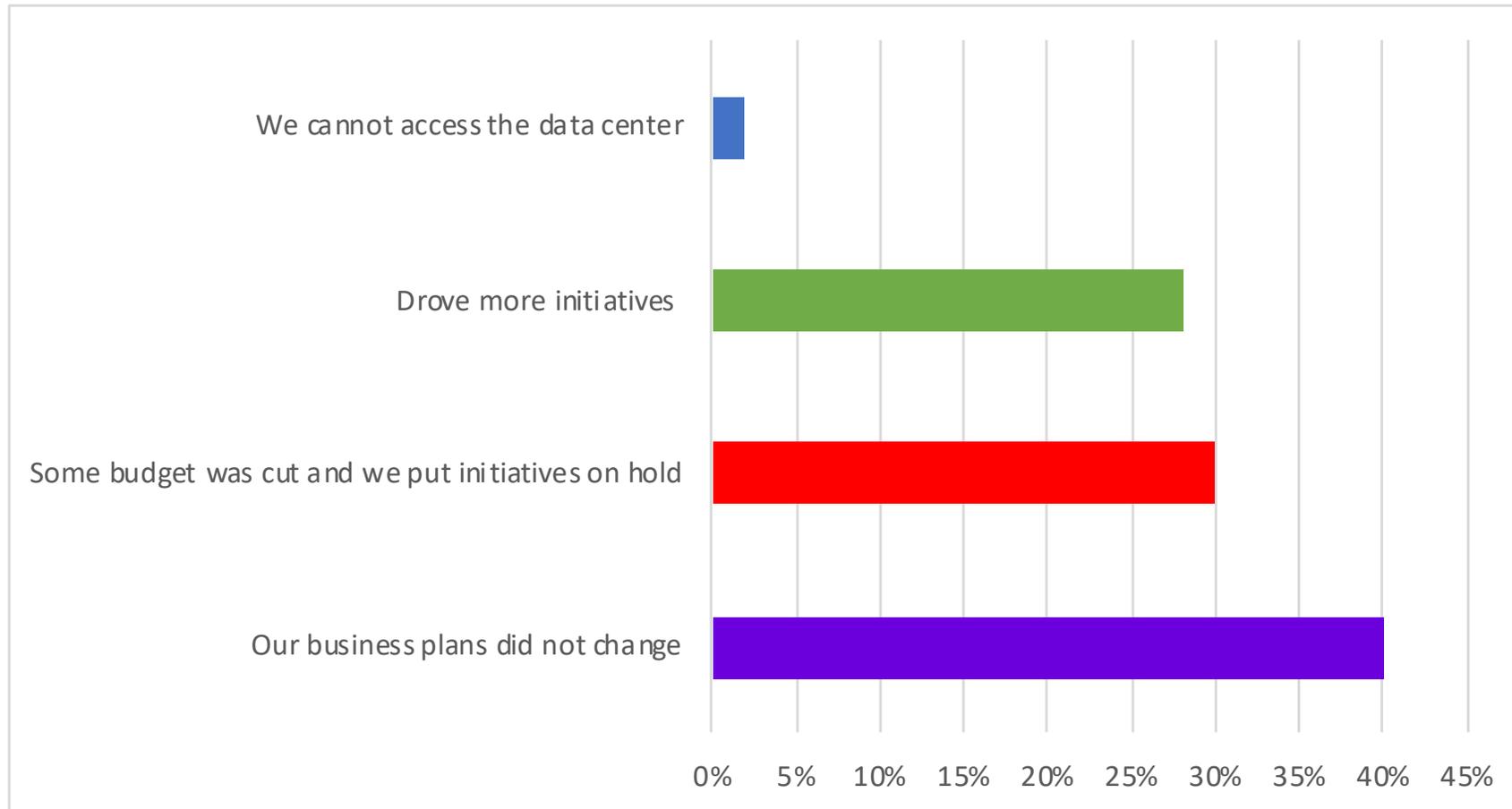
We are already seeing adoption of 200Gbps and 400Gbps. The phenomenon of adopting faster networks is consistent across geographies and verticals.

07

# COVID-19 IMPACT

# THE EFFECTS OF COVID-19 ON BUDGETS

It's interesting to see that 40% of businesses did not have any change in plans due to COVID-19. 30% did put initiatives on hold, which was almost offset by the 28% who indicated that COVID did, in fact, drive more initiatives toward the public cloud, perhaps through the inability to access the data center.

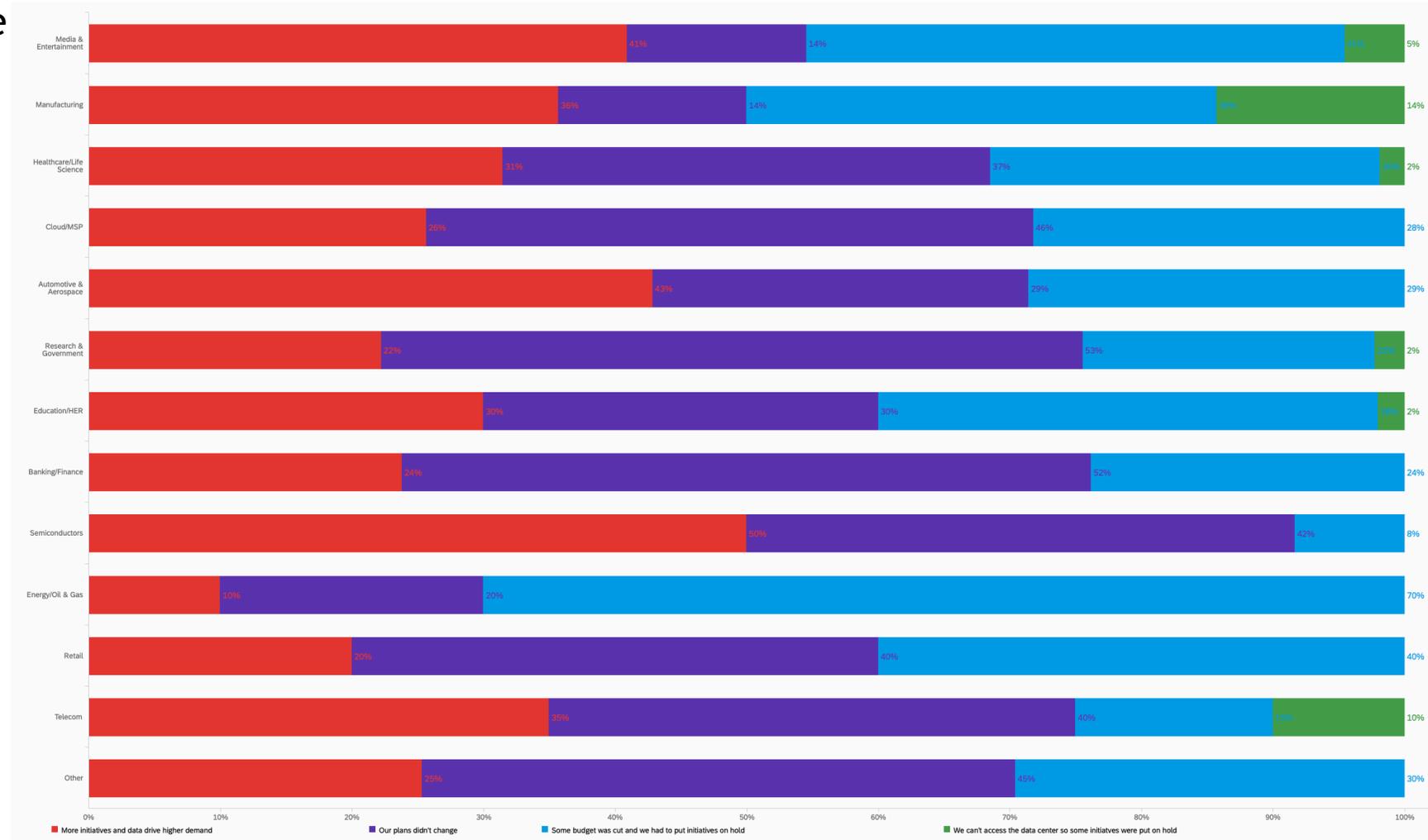


# THE EFFECTS OF COVID-19 BY VERTICAL

COVID-19 actually increased the demand for compute and usage of data in verticals such as Semiconductors, Automotive, and Media.

Oil and Gas companies seem to have cut their budget in 2020.

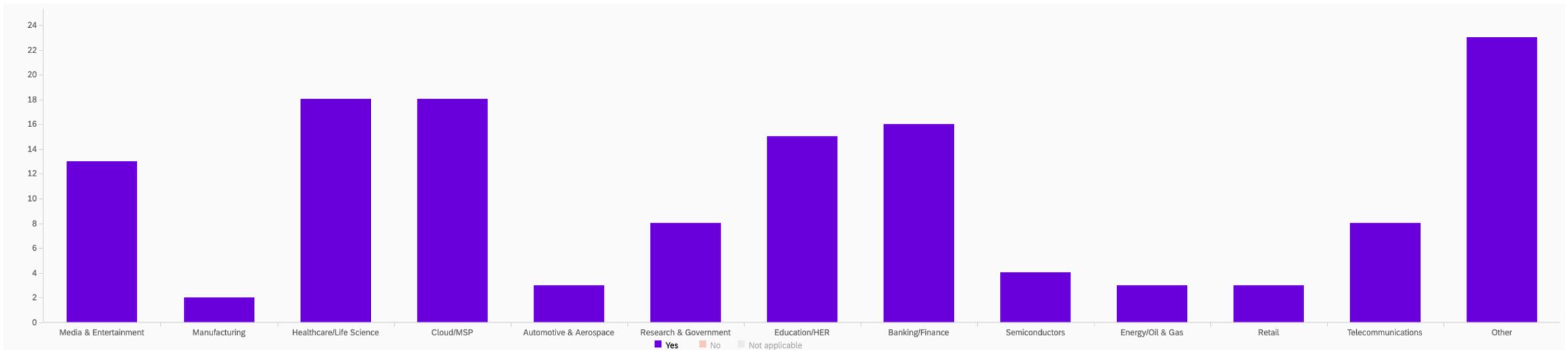
8 respondents mentioned that they no longer have access to their data center, so some initiatives were put on hold.



# DID COVID-19 ACCELERATE CLOUD ADOPTION?

We wondered if COVID-19 would have an impact on cloud adoption. We suspected that new initiatives in Life Science and Research, the availability of more data than ever, and the sudden transition to remote work would all impact the adoption of cloud. Over a third responded that cloud adoption increased due to COVID. Half, however, mentioned no increase in adoption.

The verticals that claimed to have higher adoption of cloud due to the pandemic are Life Science (no surprise here), Cloud (due to the overall increased demand for cloud instances), and Banking, which was already showing strong cloud adoption and signs of growth.

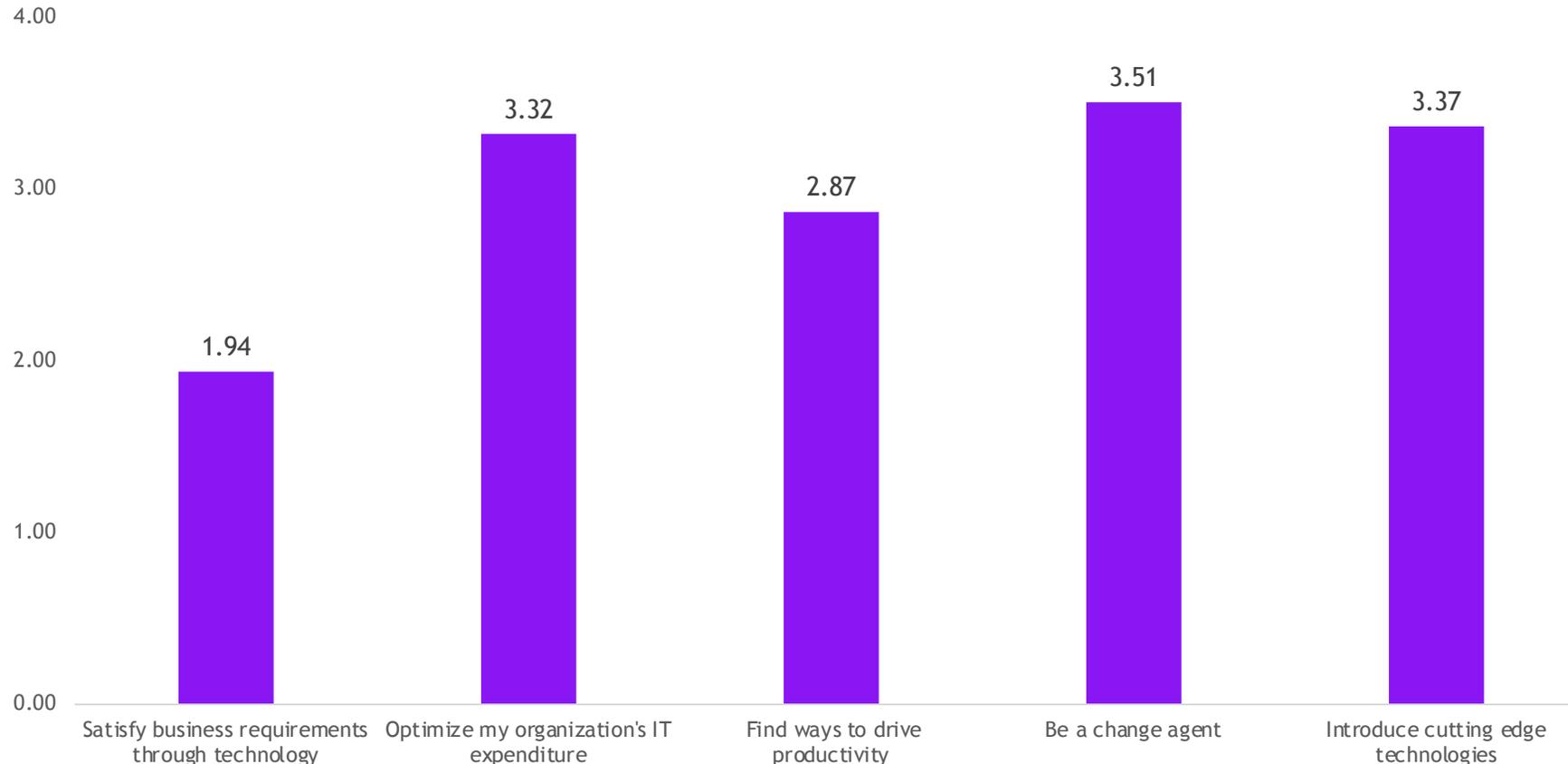


08

# ROLE AND OBJECTIVES

# WHAT ARE THE KEY PRIORITIES IN YOUR ROLE

We asked respondents to rank—on a scale of 1 to 5, with 5 being the highest priority—the key priorities in their role. Acting as a change agent seems to be a key priority, with over 200 ranking that as priority 4 or 5. Our respondents love to work on new technologies. That, together with managing and optimizing cost, came as 2<sup>nd</sup> and 3<sup>rd</sup> priorities.



# ABOUT WEKAIO

## About WekaIO

WekaIO (Weka) is used by 8 of the Fortune 50 enterprise organizations to maximize the value out of their data. Purpose-built to unlock the full capabilities for the data center of today and the future, the Weka Limitless data platform eliminates the tradeoff between Simplicity, Speed and Scale. Built on WekaFS, the Limitless data platform is optimized for NVMe-flash and the hybrid cloud so that customers can capitalize on the advances in Cloud, GPU Compute and Flash technologies without any compromises.

For more information, please visit  
<https://www.weka.io/>

## WekaIO For AI/ML/DL

Weka Data Platform is purposely built for data intensive workloads such as Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL). Its data management capabilities can accelerate time to insight/time to EPOCH by as much as 80x. Its architecture and performance is designed to maximize your usage of GPUs across cloud, on-prem or hybrid deployment.



WEKA

