



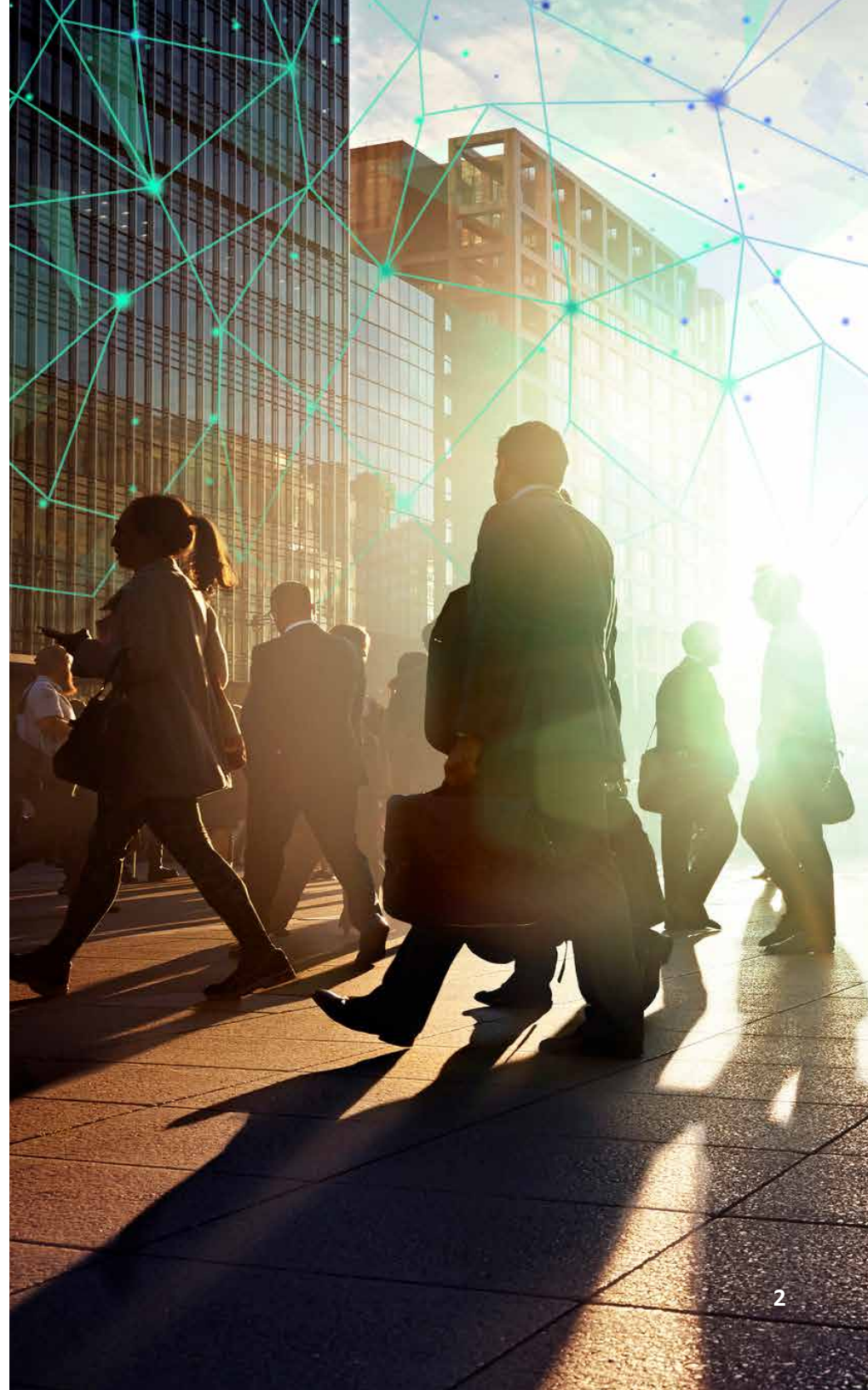
# Harness data to reinvent your organization

Build a data strategy for the next wave of cloud innovation

# Welcome to the generation of reinvention

It's hard for any organization to sustain success for a long period of time. In order to stay relevant, organizations must periodically reinvent themselves. The introduction of the cloud set off a generation of reinvention. Now, the next wave of reinvention will be driven by data. Leaders need to be able to rely on data to make informed decisions, look around corners, and take meaningful action. Building a data strategy is imperative for organizations that want to stay relevant now and in the future.

Reinvention-minded leaders need to be relentless about getting to the truth. That means having the tools it takes to pivot when needed to act on opportunities and threats. To do this, you need to become data-driven.



## Data drives the journey to reinvention

Organizations that are data-driven treat data as an asset, no longer the property of individual departments. They set up systems to collect, store, organize, and process valuable data and make it available in a secure way to the people and applications that need it.

To collect, store, organize, and act on their information, these organizations consolidate their data into centralized data lakes for easier discovery, governance, and access. They also use technologies like machine learning (ML) to unlock value from their data, such as improving operational efficiency,

optimizing processes, developing new products and revenue streams, and building better customer experiences.

For example, Moderna and other pharmaceutical companies use data and analytics to bring drugs and vaccines to market faster than ever before. BMW and many more manufacturers use data to optimize their supply chains and improve production capacity. And ENGIE, along with others in the energy sector, uses data to find new ways to reduce costs for their customers while meeting bold and aspirational sustainability goals.





## Key challenges and considerations

1

The first challenge that organizations typically face is comprehending the sheer size and scale of the data they handle every day—and the exponential growth that continues every year. In fact, more data will be created over the next 36 months than during the prior 30 years combined.<sup>1</sup> The old on-premises tools and legacy data stores from the past are not going to meet today's demands. To handle the massive scale and tremendous growth in data volumes that we see today, organizations need new data stores that can scale and grow as business needs change—whether from the gigabytes and terabytes handled today or the petabytes and exabytes we'll see managed in the future.

2

Secondly, organizations need to easily access and analyze expanded types of data, including log files, clickstream data, voice, and video. These wide-ranging data types come from a variety of sources and are stored in silos across multiple data stores. To gain valuable, new insights from all this data, organizations must break down the data silos so that their teams can access and analyze all of the relevant data regardless of where it lives.

3

The third major challenge that organizations face is adapting with greater urgency to changing customer preferences and market dynamics. To make better and faster decisions, organizations need to empower their employees with secure access to data and the ability to perform analytics and machine learning on their data in an agile and cost-effective way. Organizations running their operations on legacy on-premises data infrastructures spend a great deal of time on hardware and software installation, configuring the infrastructure



for performance and availability and spending unnecessary time on capacity planning to scale their systems. All of this unnecessary effort reduces agility and impedes quick decision-making.

4

The fourth challenge is making machine learning work. While ML is a disruptive technology that fuels innovation, organizations are struggling to make meaningful progress scaling machine learning in their businesses. According to a Gartner report, organizations with AI experience moved just 53 percent of their AI proof-of-concept pilots into production over the past two years. The lack of ML skills, organizational inertia, and quantity or quality of data to train on are just some of the issues slowing progress in this important area.

5

Finally, in a world that is increasingly dependent on data security, privacy, and compliance regulations, organizations need to be able to carefully define, monitor, and manage access to specific pieces of data through tried-and-tested data governance and security controls. They need to do this not just for the data in their individual data silos but in a comprehensive and unified way across all their data stores.

## Trends that are impacting how you get insights from data



More data than ever is being generated



Data is being stored in silos across multiple data stores



Urgency by the business to use data to make better and faster decisions



Machine learning adoption is challenged by lack of skills and organizational inertia



Data security, privacy, and compliance regulations are increasingly important



# How to become data-driven

The three stages to achieving organizational reinvention

1

**Modernize**  
Your data infrastructure

2

**Liberate**  
Put your data to work

3

**Innovate**  
Invent new experiences and reimagine old processes

# 1 | Modernize your infrastructure with the most scalable, trusted, and secure cloud provider

For organizations running legacy data infrastructures on-premises or self-managed in the cloud, the oversight of this infrastructure is tedious, time-consuming, and expensive. Concerns can arise regarding hardware and software installation, configuration, and performance and availability. Scalability requirements such as capacity planning, cluster scaling, and security and compliance issues are also concerning. Many of the on-premises data stores are remnants of old-guard commercial-grade database providers like Oracle and Microsoft with SQL Server. These infrastructures are expensive, proprietary, and often include costly vendor lock-in and punitive licensing terms.

By modernizing the data infrastructure, organizations can move from on-premises data stores to cloud data infrastructure. With AWS, organizations access IT resources, like storage, database, analytics, and machine learning, over the internet instead of buying, owning, and maintaining physical data centers and servers themselves. AWS services take care of all

management tasks such as server provisioning, patching, configuration, and backups. For example, Amazon Aurora continuously replicates six copies of the data across three Availability Zones (AZs) and transparently recovers from failures in less than 30 seconds. This lets organizations save time and costs and improve performance, availability, and scale.

When choosing the right cloud provider to trust with their data, organizations need reassurance that their choice of technology will deliver value from their data while keeping it secure and compliant across a broad and ever-changing set of regulations. They want a technology provider they can have confidence in and that understands their use cases and will grow with them as their data scales. AWS answers all those concerns with unmatched experience, scalability, reliability, security, and performance for making the most of data in the cloud.

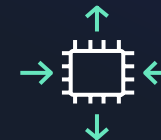
Legacy data infrastructure requires a lot of undifferentiated heavy lifting



**Hardware and software installation**



**Performance and availability**



**Capacity planning and cluster scaling**



**Security and compliance management**



**Expensive, proprietary, and high amounts of lock-in**

Internally, AWS says that there's no compression algorithm for experience, and that's because you can't learn certain lessons until you get to different milestones at scale. AWS is built for scale. Amazon Simple Storage Service (Amazon S3) provides industry-leading storage scalability, data availability, security, and performance. It is built from the ground up to deliver 99.999999999 percent (11 nines) of durability and stores data for millions of applications for companies all around the world. Hundreds of thousands of customers are already using AWS to extract value from their data. AWS has the largest and most dynamic customer base with tens of thousands of partners globally that trust AWS to run their data and machine learning workloads at scale.

AWS is also architected to be the most secure cloud-computing environment. Our core infrastructure is built to satisfy the security requirements for the military, global banks, and other high-sensitivity organizations. This is backed by a broad set of cloud security tools that ensure compliance and data governance.



*"[The AWS Cloud] is integral to our enterprise IT transformation as we look for better ways to serve our customers, streamline the way we work, and compete globally. AWS, with its proven experience and highly performant global infrastructure, will deliver the agility, speed, and scalability 3M needs to launch new business processes and service models. We look forward to expanding our use of AWS's portfolio of services, including analytics and machine learning, to gain greater insights and become an even more agile company in the cloud."*

**John Turner**

Vice President, IT Systems and Chief Information Officer (CIO), 3M





## The benefits of Amazon Relational Database Service (RDS)

IDC conducted in-depth research and found that customers who moved their databases from on-premises to Amazon RDS achieved:

**86%**

faster deployments  
of new databases

**97%**

less unplanned  
downtime

**5-month**

average investment  
payback period

## Benefits of moving to AWS for data lakes, analytics, and ML services

IDC conducted in-depth research and found that customers who moved their on-premises analytics solutions (data warehousing, data lakes, and ML) to AWS were able to realize:

**48%**

reduced  
cost of ownership

**415%**

five-year ROI

**76%**

reduction in  
unplanned downtimes



Best Western experienced tremendous success migrating its data from Oracle to Amazon Aurora, our MySQL- and PostgreSQL-compatible relational database built for the cloud. The hotel company can now process 2.3 billion data points in a single month in less than a minute. The migration also saved them over half a million dollars in hardware and software costs, as well as future savings.

*"The hotel industry is rapidly changing, as more customers expect the ease and convenience of mobile computing. Moving to AWS brings our organization to the forefront of innovation and allows us to give our guests fast, reliable, and secure data processing so they can organize their trips, change their reservations, and book their stay with us."*

**David Kong**  
President & CEO, Best Western



## SAMSUNG

### Samsung migrates 1.1 billion users across three continents

After migrating 75 petabytes of data to cost-effective, purpose-built, and time-saving AWS database services, Samsung turned off its final Oracle database, both modernizing its data infrastructure and increasing data access, consistency, and performance as a whole. Compared to its previous Oracle solution, the company saved 44 percent on monthly operational costs and another 22 percent in maintenance fees by using Amazon Aurora.



### Nasdaq scales with billions of records effortlessly

Over the last decade, Nasdaq has moved much of its U.S. market's transactional data from a legacy on-premises data warehouse to Amazon Redshift and Amazon S3 as its data lake to meet the increased data demand of its high-volume electronic markets.

As volatility flexed markets, Nasdaq was well-positioned to jump from processing 30 billion records to 70 billion records per day and even 115 billion records at its peak.

## Blackboard

### Blackboard modernizes to eliminate licensing costs

For Blackboard, educating millions of people worldwide meant paying pricey licensing costs. The company eliminated that overhead by migrating its database and modernizing its infrastructure with the open-source PostgreSQL database on Amazon RDS. The migration involved moving thousands of virtual machines and multiple petabytes of data to AWS. It also freed transactional data without increasing computational workloads—thereby allowing Blackboard to shut down over 20 data centers and rely on just nine core global data centers.

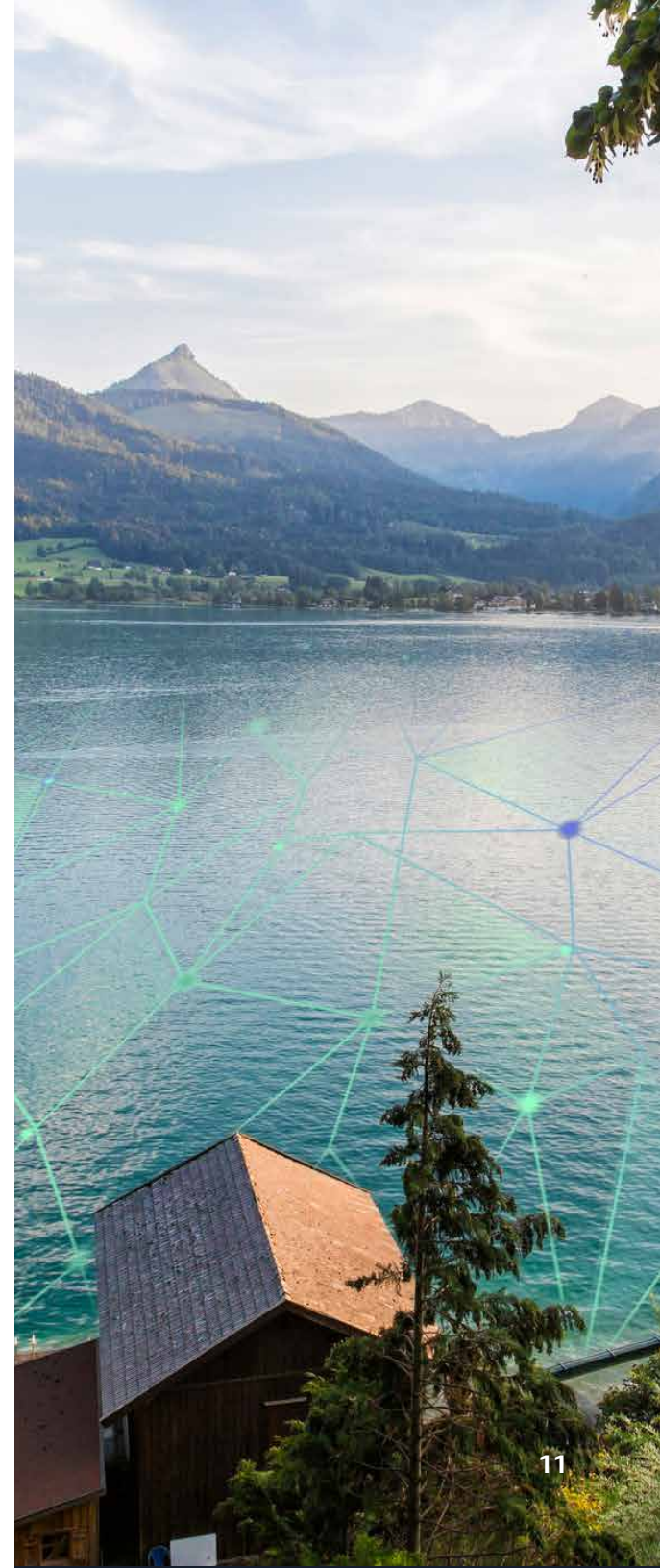
## 2 Put your data to work: Data lakes and purpose-built data stores

To make decisions quickly, organizations need to store any amount of data in open formats and break down disconnected data silos. Their employees need to be empowered to run analytics or machine learning using their preferred tools or techniques and manage user access to specific pieces of data with the proper security and data governance controls. AWS helps organizations do all of this through our Lake House approach, which brings together the best of data lakes and purpose-built data stores.

With a Lake House approach, organizations can move any amount of data from various silos into an Amazon Simple Storage Service (Amazon S3) data lake. Unlike other cloud providers, organizations can store their data in S3 using standards-based open data formats to avoid being locked into any one proprietary data format or approach to analytics. Storing data in standards-based open formats makes it easy for any analytics or machine learning service to work with the data. It also eliminates the need to unnecessarily move, transform, or reformat the data in order to gain value from it. This is

particularly useful when working with petabyte- and exabyte-scale data. For example, Amazon Athena is an interactive query service that lets organizations instantly analyze data stored in S3 using standard SQL without having to set up and manage any servers.

In addition to using a data lake, organizations also use purpose-built data stores to get the best performance, scale, and cost advantages for their specific use cases. Amazon Aurora processes transactions at high speed, and Amazon Elasticsearch Service stores and analyzes large volumes of log data at low cost. In a world in which organizations are increasingly required to work with terabytes, petabytes, and even exabytes of data, purpose-built data stores have the ability to run a particular workload or use case extremely well. All of which is why AWS spent the last several years building the right tools for the right job: 15 purpose-built databases, 12 purpose-built analytics services, and 30 machine learning services. That is more than you'll find anywhere else by a fair amount.





As more and more organizations store their data in S3 data lakes and also in purpose-built data stores, they need to frequently move their data back and forth between their data lakes, data warehouses, and purpose-built stores. Amazon Redshift and Amazon Athena both support federated queries and the ability to run queries across data stored in operational databases, data warehouses, and data lakes. Federated queries can provide insights across multiple data sources with no data movement and no need to set up and maintain complex extract, transform, and load (ETL) pipelines. Amazon Redshift data lake export allows organizations to unload data from their data warehouse to their data lake in open formats, ready for analytics. With the Lake House approach, organizations can also use capabilities like AWS Glue Elastic Views to effortlessly move and sync data between data lakes, data warehouses, and purpose-built stores. This gives them the scale and flexibility of storing and processing their data in a data lake, with the performance and cost-effectiveness of using purpose-built data stores.

Finally, the Lake House approach empowers developers, business analysts, and data scientists to break down silos, and discover, collect, and analyze data in a secure and governed way. The approach provides organizations with capabilities like AWS Lake Formation, which includes a Data Catalog that automatically discovers, tags, and catalogs data. It sets up an easy way to centrally define and manage security, governance, and auditing policies—all in one place. This enables organizations to provide fine-grained access of data to the right user at the right time, which in turn effectively meets their regulatory governance and compliance requirements.

## Lake House approach on AWS





### The American Heart Association uncovers lifesaving insights

The Precision Medicine Platform developed on the AWS Cloud by the American Heart Association allows researchers to uncover critical cardiovascular insights via easily searched and accessed centrally stored data.

*“By using the AWS Cloud, the Platform will harness the power of big data to revolutionize the way cardiovascular research is performed and speed the promise of precision cardiovascular medicine.”*

**Nancy Brown**

Chief Executive Officer, American Heart Association



### Moderna accelerates mRNA COVID-19 vaccine development

The scalable, high-performance computing capabilities of AWS data storage enable Moderna Therapeutics to run its Drug Design Studio. It successfully converts analytics and machine learning findings into sequences, producing one of the first COVID-19 vaccines to gain approval.

*“With AWS, our researchers have the ability to quickly design and execute research experiments and rapidly uncover new insights to get potentially life-saving treatments into production faster.”*

**Stéphane Bancel**

Chief Executive Officer, Moderna



### Epic Games designs megahit *Fortnite* around player behavior

Epic Games' designers gain deeper insights into game-player behavior with the readily accessible data sent to and from AWS via its interactive, highly popular gaming experiences, including the megahit *Fortnite*. Leveraging the power of the AWS Cloud, *Fortnite* designers can create constant feedback loops that help increase customer satisfaction in real time.

### 3 Invent new experiences and reimagine old processes with machine learning and AI

Machine learning is one of the most disruptive technologies of our generation. It can help increase revenue opportunities, inform better and faster decisions, and improve operational efficiencies. In the fullness of time, virtually every application will be infused with machine learning and artificial intelligence. AWS meets customers wherever they are on their ML and AI journey, helping them achieve their unique business outcomes. Builders of all levels of expertise can access the broadest and most complete set of ML and AI services from AWS.

The end goal of becoming data-driven is to build the capabilities necessary to reinvent how your teams deliver value to users, customers, and the world using your data. ML- and AI-powered innovations are the instrumental components of this type of transformation across and within industries.

For expert practitioners, AWS supports all the major machine learning frameworks, including TensorFlow, MXNet, PyTorch, Caffe 2, etc. AWS offers the highest-performance instances for ML training in the cloud with Amazon EC2 P4d instances, powered by the latest NVIDIA A100

Tensor Core GPUs and coupled with first-in-the-cloud 400 Gbps instance networking. P4d instances are deployed in hyper-scale clusters (called EC2 UltraClusters), offering supercomputer-class performance for the most complex ML training jobs. For inference—representing 90 percent of ML costs—Amazon EC2 Inf1 instances powered by AWS Inferentia chips are the most affordable in the cloud.

For data scientists and ML developers, AWS offers Amazon SageMaker, the industry's most comprehensive, managed machine learning service. It was built from the ground up to simplify the process of machine learning with tools for every step of ML development. Those tools automate the jobs of labeling, data preparation, feature engineering, statistical bias detection, AutoML, training, tuning, hosting, explainability, monitoring, and workflows.

By standardizing on SageMaker, teams can remove the complexity from each step of the ML workflow to prepare, build, train, and deploy high-quality ML models more quickly and cost-effectively. The efficiency benefits are potentially game-changing. SageMaker-



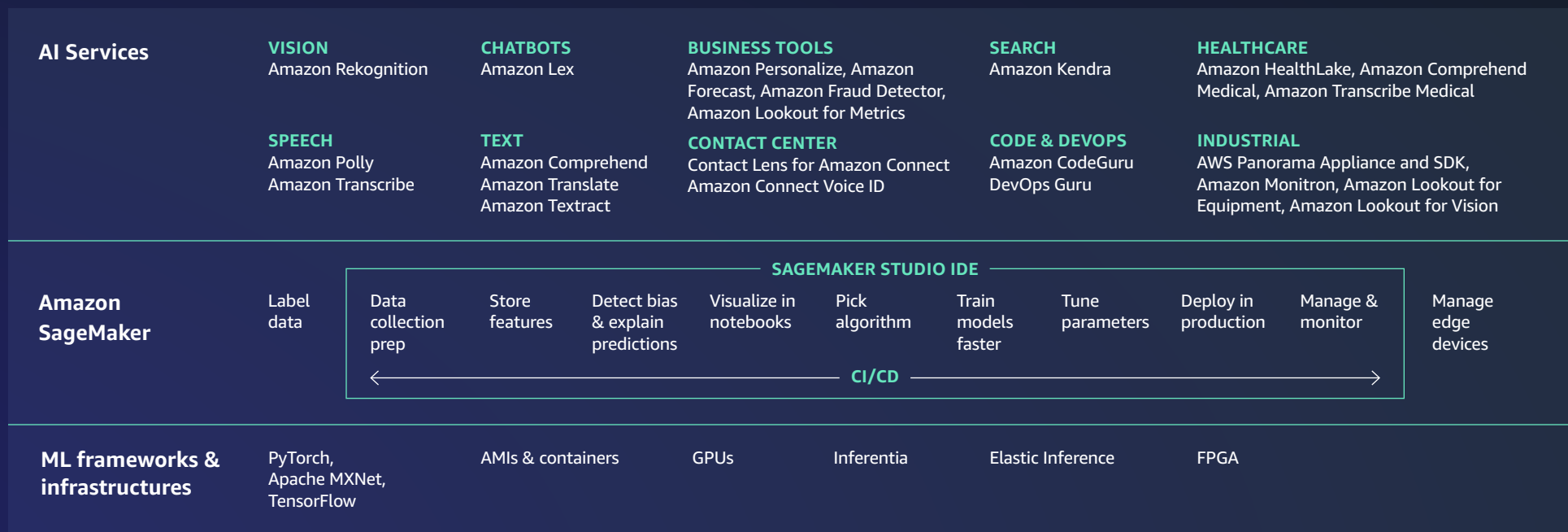
equipped data scientists are up to 10 times more productive in preparing, training, and deploying high-quality machine learning models.

For developers and business users, AWS offers pre-trained AI services that provide ready-made intelligence for applications and workflows. Utilizing AutoML technology, these end-to-end services are built to solve business needs right out of the box. They address common use cases such as personalized recommendations, contact-center intelligence, document

processing, intelligent search, business-metrics analysis, and more. AWS also offers industry-specific AI services for both industrial and healthcare industries.

For machine learning to be used more widely, it needs to be brought closer to the data lakes and purpose-built data stores, where much of the data needed for machine learning resides. To do that, AWS provides built-in integration of machine learning as part of its purpose-built data stores and business intelligence (BI) services. Developers can use Amazon Aurora ML to

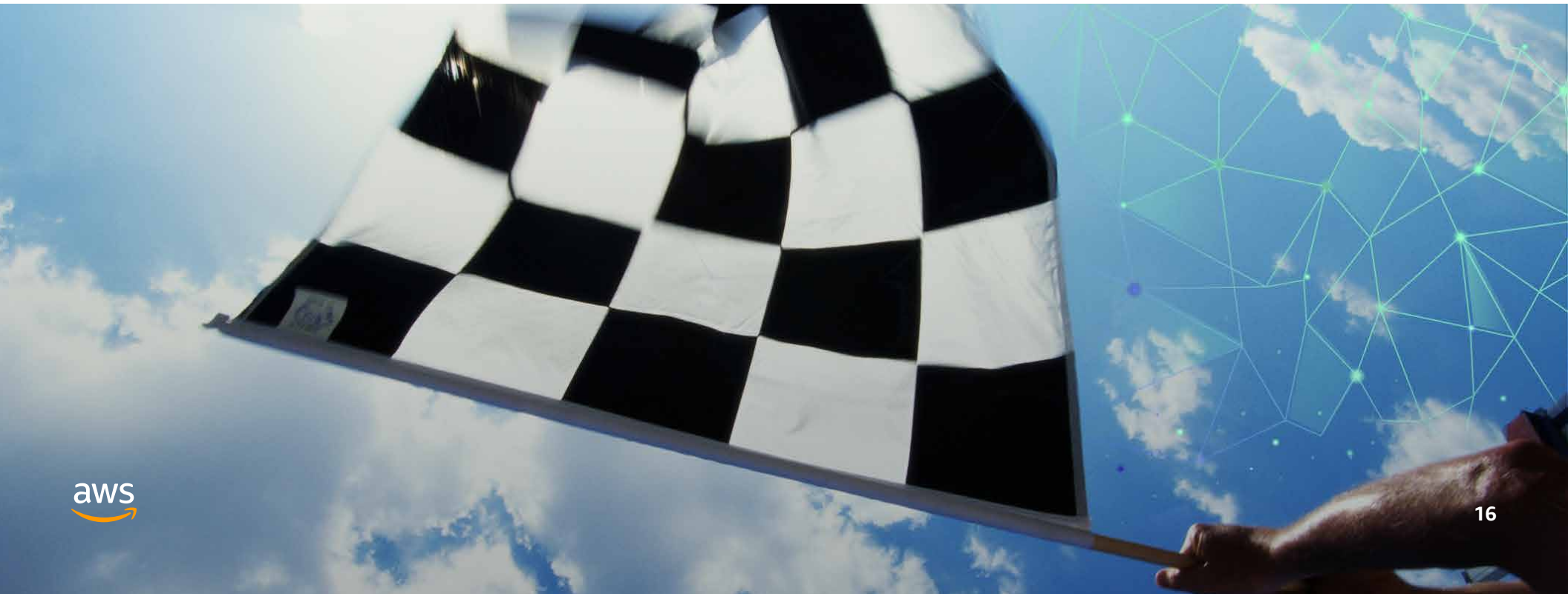
## Invent new experiences and reimagine old processes with machine learning and AI that match your business needs



run machine learning with a simple SQL query on transactional data or use Amazon Neptune ML to apply deep learning to graph data without having to build and train machine learning models. Likewise, data analysts can use Amazon Redshift ML and Amazon Athena ML to run machine learning on their data in a data warehouse or data lake without having to select, build, or train ML models. And, business analysts can utilize Amazon QuickSight Q, which employs machine learning to automatically generate a data model that understands the meanings of (and relationships between) business data—asking questions of the data using plain language and receiving answers in near real time.

In addition to technology, AWS offers several services and related features to help organizations get started. These services help teams overcome the

challenges of implementing these technologies, often revolving around data ambiguity, uncertain costs, lack of necessary skills, and overall complexity. The good news is that organizations can accelerate their machine learning projects with access to fun, hands-on learning tools. For example, DeepRacer is a fully autonomous 1/18th scale race car driven by reinforcement learning. Developers can compete in the DeepRacer League to gain and demonstrate usable skills, and enterprises can launch their own leagues to train internal developers. Over 150 global organizations—including Capital One, Moody's, Accenture, DBS Bank, BMW, and Toyota—have trained thousands of developers with DeepRacer enterprise events. AWS also offers the experts at Machine Learning Solution Lab, broad and custom ML training, and a network of 70+ partners to help organizations get started on the machine learning journey.





## NFL player-tracking excites a new generation of fans

AWS and Amazon SageMaker power pro football innovations with RFID chips. These chips are embedded in NFL player shoulder pads to drive new insights and generate predictions on which gameplays and content will excite audiences.

*"I've gotten a lot of positive feedback from fans saying, 'Wow, how did they complete that pass?' We've been able to quantify it and compare it to other passes, and that's been a real value-add for fans because it creates context for what's happening in the game."*

Matt Swensson, Vice President of Emerging Products and Technology, NFL

## PHILIPS

### Philips improves medical decision-making and patient outcomes

Philips uses AWS for the development of data, analytics, ML, and computer vision technologies that enable healthcare providers to make better decisions for their patients. The company uses AWS for its HealthSuite Digital Platform, which securely consolidates information from patient records, wearables, home-based remote monitoring equipment, insurance companies, and healthcare organizations.



Fannie Mae®

## Fannie Mae infuses machine learning into financial services

Like all large financial services companies, Fannie Mae absorbs tremendous amounts of data. Just in the process of assessing property values, the financial leader receives close to 40,000 appraisal reports and over half a million images daily. Fannie Mae stores its data in AWS Cloud and utilizes Amazon SageMaker to develop, evaluate, expand, and improve the models used in everything from assessing the loans it backs to analyzing property values.

*"We're ingesting new data sources every week, every month, to look for new insights. Applying that data to help lenders make the right decisions requires the right tools."*

Scott Hallworth, Senior Vice President and Chief Data, Modeling, and Analytics Officer, Fannie Mae



## Reinvent your working culture around data

Being data-driven requires a cultural change in which every business goal and decision is supported by data. Beyond the big decisions, data should be put in the hands of everyone for everyday decision-making. It requires a culture of experimentation where failures are expected, and leaders take stake in the success of ongoing data initiatives. This is about democratizing action, not only access. It is about building data literacy, designing education plans for the unique needs of various roles and skill sets across the organization, and making the necessary tools continuously accessible to everyone who needs them.

To start, look for silos and eliminate them along with barriers, resistance, and the inertia of returning to old ways. Bring technology and domain experts together to solve the right business problems with data and develop the shared buy-in that encourages long-term adoption.

### Common traits of data-driven organizations

- Everyone has access to data, starting from the top
- Organizational capabilities support data-driven culture
- Experimentation creates organizational improvements
- Analytics back transformation with AI and ML
- Silos break down while data transparency increases

## Five steps to reinvention as a data-driven organization

### Step 1

Investigate how data flows in your organization and what gatekeeping controls are in place. Uncover data silos and gauge the level of difficulty for employees to access the data they need.

### Step 2

Ensure a senior, well-respected, and empowered leader is driving the cultural initiative to become truly data-driven.

### Step 3

Treat data as a product, in part, by bringing application engineers and data engineers together. Closely align data, product, and integration strategies.

### Step 4

Make IT a key player. IT has a unique view of the end-to-end business cycle, cross-departmental workflows, and transactional systems that hold valuable insights.

### Step 5

Create a data governance structure that enables employees rather than restricting them.

## Pinterest reinvents as data-driven

*"Pinterest is data-driven, top to bottom, in everything we do—but we never lose our focus on the individual people who use our product."*

David Chaiken, Chief Architect, Pinterest

Founded in 2010, the visual discovery engine Pinterest supports 400 million users each month. At Pinterest, users known as "Pinners" log more than 240 billion information points known as "pins." With all that data, technology plays an integral role in supporting the company mission: Provide value to all who use Pinterest for pinning everything from recipes to craft ideas, travel, and more.

In fact, almost all changes to Pinterest products are data-based and come from research done on Pinner behavior. Using data and ML, the company collects and analyzes trending searches and overlapping interests on a millisecond-by-millisecond basis to identify relevant ideas for each person. Pinterest relies on data to analyze what it should do next and how its Pinners respond to changes and improvements. It constantly tracks how and where

it can best serve users. And although the provider started as a website, it's evolved to deliver mobile experiences to 70 percent of Pinners who log in via their mobile devices.

AWS has long supported Pinterest's remarkable growth, providing the scalability and reliability that it must have to run its business. Chaiken says, "As a born-in-the-cloud company," Pinterest's ambitions "have never been limited by the walls of a data center." AWS has helped eliminate data constraints for Pinterest and thousands of other organizations that have endeavored to build their functions (and futures) around data.

The next wave of reinvention will be driven by data. Leaders and other decision-makers looking to join that wave need to be tenacious about getting to the truth. They also need the essential tools to stay agile enough to pivot when needed to act on new opportunities. Simply stated, you need to become data-driven. Organizations that are data-driven seek the truth by treating data as an organizational asset, no longer the property of individual departments.

Today, hundreds of thousands of organizations rely on AWS to reinvent their operations and become data-driven. They choose AWS for data and machine learning to:

- Modernize their data infrastructures with the most scalable, trusted, and secure cloud provider
- Put their data to work with the best of both data lakes and purpose-built data stores
- Invent new experiences that match evolving business needs and reimagine old processes with machine learning and AI

AWS provides the hands-on experience, purpose-built tools, trusted data infrastructure, and proven partner ecosystem to help you on your journey to reinventing your organization with data.

**[Learn more about reinventing as data-driven »](#)**