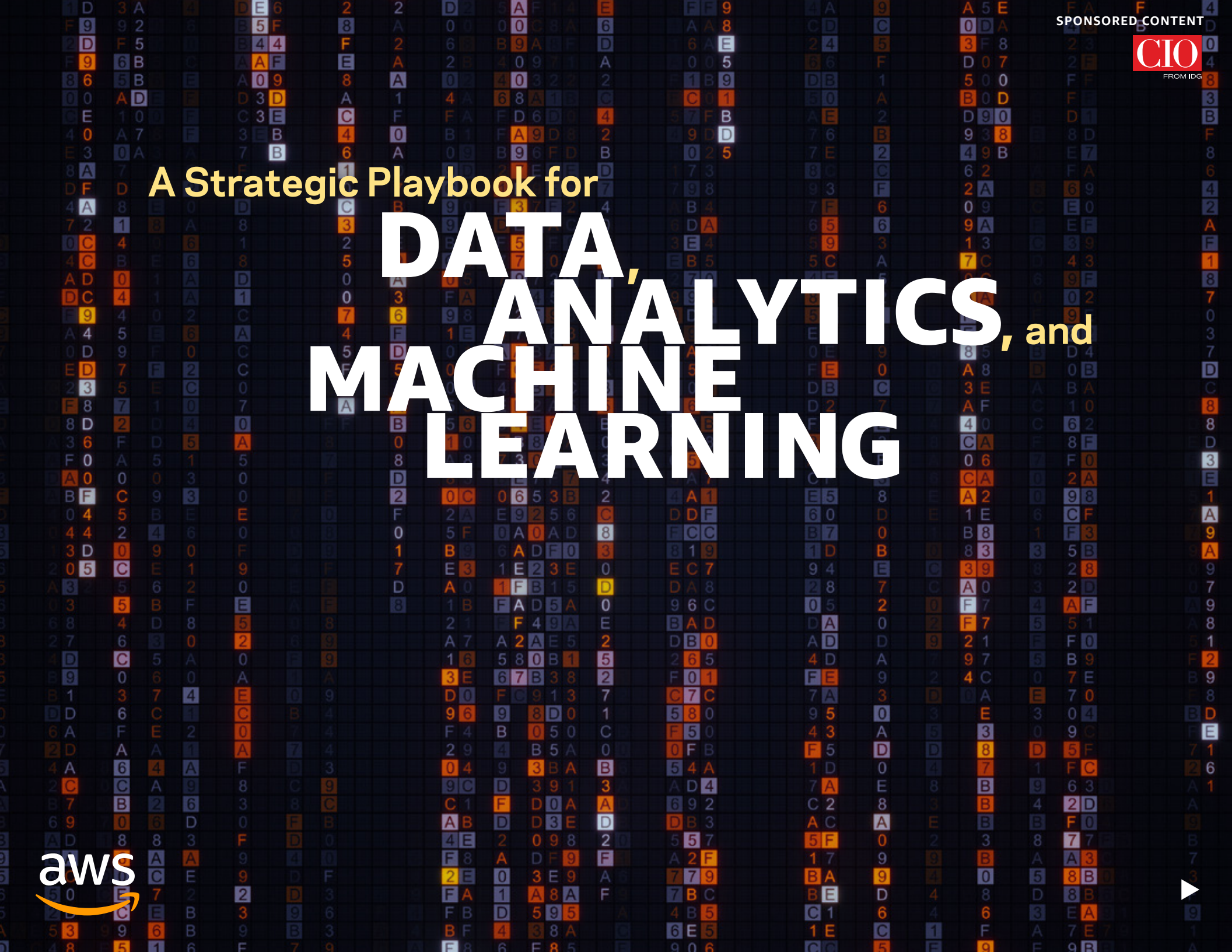
CIO
FROM IDG



A Strategic Playbook for

DATA, ANALYTICS, and MACHINE LEARNING

aws

Introduction

Organizations that once stored gigabytes of business data now find themselves swimming in petabytes or even exabytes of information. Harnessing the value of these massive data volumes requires a modern, cloud-based data infrastructure that unifies disparate information silos. This approach creates a universal version of "truth" and empowers everyone in an organization to make more informed decisions and act with confidence.

A few key trends have brought us to this point:

- **The digitalization of businesses,** coupled with a borderless Internet that extends application reach anywhere, has given rise to unprecedented data growth from mobile devices, IoT devices and sensors, cloud-based apps, social media, log files, click-streams, and other sources.
- **The cloud has driven down** the cost of storage, allowing organizations to cost-effectively manage the deluge without having to decide up front what data is worth keeping and what should be discarded.
- **The cloud is also the conduit** to highly accessible, pay-as-you-go compute and storage resources, which makes it easy for organizations to scale based on need while harnessing and analyzing every last bit of data for critical business insights.

Organizations now have the opportunity to marshal data, analytics, and machine learning (ML) to promote critical business decision-making. Such data-driven insights can guide countless business scenarios, including:

- **Launching** new product offerings
- **Innovating** on new business models or revenue streams
- **Automating** manual processes to improve efficiency
- **Improving** customer experiences to build trust and loyalty
- **Optimizing** interactions with business partners

Regardless of where you are on your data modernization journey, this playbook will help you refine your strategy to effectively scale data, analytics, and machine learning across the enterprise, so you can accelerate innovation and drive your business forward.

Contents

Wanted: A modern, data-driven enterprise

Becoming a data-driven organization is achievable through a mix of technology, people, and processes, bolstered by a broad view of data as a strategic asset. While most organizations have data-driven transformation in their sights, the shift remains elusive for many:

- **99%** of organizations have invested in data and artificial intelligence (AI) initiatives
- **65%** have spent more than $50 million on these efforts

But…

- Just **38%** say they are a data-driven organization
- Only **27%** say they've successfully built a data-driven culture (*Source*: 2020 NewVantage Partners Big Data and AI Executive Survey)

The numbers show a clear gap between organizations' aspirations for using data strategically and the reality of what they've accomplished to date.

Some have already taken steps to close the gap, using the cloud as a foundation for a modern data architecture:

- Georgia-Pacific migrated 50 TB of structured and unstructured production data from a legacy database infrastructure to a cloud-based data lake in order to cost-effectively ingest, transform, store, and analyze that data. Layering analytics tools on the data has helped Georgia-Pacific optimize key manufacturing processes, including the ability to predict equipment failure 60 to 90 days in advance, which reduces unplanned downtime.
- Zappos is using cloud-based analytics and machine learning to personalize product sizing and search results for retail shoppers while preserving a highly fluid and responsive user experience. The result: Zappos has reduced repeated searches and product returns, achieved higher search-to-product-click-through rates, and raised the position of customer selections in search results.
- After The Pokémon Company International (TPCi) launched its Pokémon GO mobile game in 2016, the number of users requiring access to the system increased to more than 300 million in two years. To address the massive influx of Pokémon GO users, TPCi migrated from a third-party NoSQL document database to AWS (Amazon Web Services) fully managed database services, which allowed it to reduce the number of database nodes from 300 to 30 and decrease its monthly database costs by tens of thousands of dollars.

These examples underscore the benefits of moving away from antiquated, monolithic applications that run on one-size-fits-all relational databases to highly distributed, microservice-based systems running on multiple purpose-built databases. It also means moving from on-premises and old-guard legacy data warehouses to open and flexible data lakes and "lake house" architectures.

"It's not a one-size-fits-all world anymore," says Shawn Bice, Vice President, Databases, with AWS. "How you approach data in the cloud is the foundation for future business growth. A native cloud application architecture allows businesses to innovate faster than ever before, at lower cost, and faster time to market because you're not limited to what one thing can do."

Limitations of monolithic architectures

Organizations that continue to struggle with legacy infrastructure likely face one or more well-known challenges involving people, process, and technology. Becoming a data-driven organization requires new thinking about all three.

Traditional data analytics architectures rooted in structured databases and relational data warehouses have grown costly to manage, update, and secure. These limitations prevent enterprises from fully monetizing valuable data. Legacy architecture wasn't built to support enterprise-wide data management and analytics at the scale required to handle petabytes or even exabytes of data across the enterprise.

Most CIOs recognize the limits of legacy data infrastructure—but may hesitate to push their organization out of its technological comfort zone. Such a short-term perspective could inhibit long-term business growth.

"Don't let familiarity turn into a blind spot that stifles innovation," says Bice.

INNOVATION SPOTLIGHT

Amazon.com

Amazon.com once maintained one of the biggest Oracle data warehouses in the world, but it wasn't enough to keep pace with Amazon's growth.

"Five years ago, our ability to grow and analyze our business was limited by technology choice," says Jeff Carter, Vice President of Data, Analytics, and Machine Learning at Amazon.

Amazon made the strategic decision to move off the legacy data warehouse and onto a cloud-native data lake architecture comprising a variety of AWS services. The migration involved moving 50 petabytes of data, 75,000-plus data warehouse tables, and 7,500 OLTP databases supporting the company's business-critical ordering, processing, and fulfillment systems.

Carter admits to some initial concerns about such a massive migration causing disruptions. "We didn't want to be the team that shut down the business," he says. "But what we found was in pretty much every instance, availability was better."

"By migrating to the AWS technologies and implementing the data lake, we have been able to scale to meet our business needs," says Carter, adding that the new environment is about 30-50% less expensive to maintain than the previous architecture.

The benefits of a cloud-based data foundation

A modern, cloud-based data infrastructure is essential to gain the flexibility and scalability needed to react quickly to changing business needs. With a modern data architecture, there are virtually no limits to how much and what kind of data an organization can store and manage, opening up myriad possibilities for leveraging information in new and better ways across the business.

"Decades ago, databases were designed to be optimized for storage because the costs were so high," explains Herain Oberoi, Director of Product Marketing, Databases, Analytics, and Blockchain, AWS. "Cloud economics have removed the constraints of having to decide what data to store and what to discard. Now the default is, let's just store everything because we might not know what we want to do with the data, what questions we want to ask, or what insights we might get down the road."

Unlike the rows/columns/table structure of a traditional data warehouse, a modern data architecture can store all types of semi-structured or unstructured data such as web logs and images, eliminating the need for separate data silos. A centralized data lake also enables the ability to tag and catalog data to make it discoverable, ingest and process data in real time via streaming technologies, and apply security controls and permissions to the data to promote governance and maintain compliance.

Elasticity is another advantage of a modern cloud-based data infrastructure. A traditional environment requires investment in software licenses, infrastructure, and data center horsepower to drive big analytics workloads, even if they are temporary to accommodate a surge or a specific use case. Not so in the cloud, where you can spin up or scale back infrastructure based on the needs of the analytics workload. That provides an opportunity to start small and go big or front-load horsepower to handle a temporary surge period and only pay for the capacity used.

"Fundamentally, cloud is about low-cost and elastic storage and compute," says Oberoi. "That makes analytics the perfect workload for cloud."

INNOVATION SPOTLIGHT

Equinox Media

As COVID-19 bore down in the spring of 2020, fitness brand Equinox recognized the need to fast-track the launch of its Variis online fitness streaming platform. With the company anticipating a surge in demand for in-home fitness services, the development team looked to learnings from Equinox's Fitness environment to develop a modern and scalable data infrastructure based on an Amazon S3 data lake, and serverless architecture including AWS Lambda, Amazon DynamoDB, and Amazon Athena.

By developing Variis from the ground up with a cloud-native infrastructure, Equinox Media was able to accelerate the new platform's launch and quickly scale up to meet demand, says Elliott Cordo, Equinox Media's Vice President of Technology Insights. Leveraging the highly scalable, serverless data platform, "we ramped from beta users to launch in just a matter of weeks," says Cordo.

"We launched a startup at very low cost with the confidence that we could scale it with reliable cost prediction," he says. "The ability to handle explosive growth clearly demonstrates the advantages of modern data engineering and cloud-native design."

How to democratize your data

Modern data management and analytics tools, combined with cloud-based infrastructure, provide the underpinnings of a unified, holistic view of data across the enterprise. But tools are only as good as the people who use them. What most often trips up data-driven initiatives are corporate culture challenges and low levels of "data literacy," as New Vantage Partners and other researchers have noted.

A critical step toward becoming data-driven involves better education and understanding of the value of data, from the CEO on down. The goal is to put data in the hands of everyone and enable them to use data for everyday choices, not just big decisions.

Ishit Vachhrajani, Enterprise Strategist at AWS and the former Global CTO at A+E Networks, prefers the term "data proficiency" to data literacy, because the goal goes beyond data awareness to knowing how to put it to use. "It's not just about democratizing access to data," he says. "It's about democratizing action using that data."

Vachhrajani recommends the following steps to create a data-proficient culture:

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

5 characteristics of a data-driven organization

1 **Go beyond** "executive sponsorship" to "executive engagement" by making visible changes, starting at the top, to put data to use across the enterprise.

2 **Focus on creating** a data-driven culture and building organizational capabilities to support the culture, at a scale beyond just a few islands of excellence.

3 **Embrace data-driven** experimentation to test many ideas and continually improve the business.

4 **Use a strong foundation** of data analytics to power transformation using artificial intelligence and machine learning.

5 **Eliminate silos** by treating data as an organizational asset, not a departmental property, and by bringing data transparency and accountability to the entire enterprise.

- **Make sure a senior,** well-respected, and empowered leader is driving the cultural initiative to become a truly data-driven company.
- **Treat data as a product,** in part by bringing application engineers and data engineers together, and also by very closely aligning data strategy with product and integration strategy.
- **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 useful insights.
- **Create a data governance structure** that enables employees rather than restricts them.

Cultural change, of course, takes time. But there are plenty of benefits to capture along the way. "The steps you take will start to deliver wins early on," says Vachhrajani. "And you'll be able to show people how the wins fit in the context of the bigger picture."

How do you know when cultural change is taking hold? When people start pushing back. "Change is uncomfortable and can cause a lot of friction," Vachhrajani says. "The stronger that resistance is, especially early on, is a sign that things are working. It's important to feel that valley of despair when you're going through change because that tells you you're moving the needle."

INNOVATION SPOTLIGHT

Fortnite

Epic Games uses AWS to deliver Fortnite to more than 250 million players around the world, along with more than 10 million concurrent players during in-game events such as the world's largest concert.

The Fortnite team needed a way to process and analyze over 100 petabytes of data (125 million events per minute) ingested from game clients and game servers to understand and adapt to player engagement. Epic Games turned to AWS for an Amazon S3 data lake in combination with Amazon EMR, Amazon EC2, and Amazon Kinesis to perform real-time analytics that fuel continuous game improvement.

Get more from your data with machine learning

With the proliferation of data and virtually unlimited quantities of computing power available via the cloud, AI and machine learning are poised to have a profound impact on an increasingly broad canvas of business and technology.

The ability to analyze massive amounts of data to derive meaningful insights can help businesses make faster, more informed decisions. It's clear that AI and ML are positioned to give companies a competitive edge:

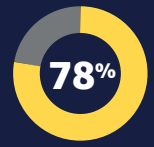
- **78%** of organizations are considering or have already deployed machine learning technologies as part of a digital business strategy. (*Source:* IDG 2019 Digital Business Study)
- **63%** of companies have invested in machine learning to close competitive gaps with rivals. (*Source:* Deloitte, State of AI in the Enterprise report, 2018)
- **54%** of businesses deploying AI have experienced productivity increases. (*Source::* PwC, AI Predictions for 2018)

Leadership teams should not limit their thinking around potential use cases for AI and machine learning. "Every part of the organization should be thinking about machine learning," says Sri Elaprolu, Senior Leader with the Amazon Machine Learning Solutions Lab.

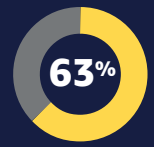
For example:

- **In sales and marketing**, machine learning and AI can help analyze activities and behavioral data to create personalized experiences for customers.
- **In finance,** the technologies are empowering more accurate forecasting models and fraud detection analysis.
- **Operations and logistics** are tapping AI/ML to optimize order fulfillment and improve shipping routes for last-mile delivery.
- **IT is parlaying tools** to aid in security threat detection and mitigation and to streamline software development operations.

Machine Learning and AI



of organizations are considering or have already deployed machine learning technologies as part of a digital business strategy



of companies have invested in machine learning to close competitive gaps with rivals

of businesses deploying AI have experienced productivity increases

"The cloud enables extremely low-cost compute and storage, which opens up opportunities for more modeling," says Elaprolu. "There's lots of innovation yet to happen. We are barely scratching the surface."

Beyond functional use cases, many sector-specific examples of AI/ML are emerging as well.

- **In healthcare,** technology company Cerner employs machine learning to help predict congestive heart failure up to 15 months before it manifests in clinical tests, improving patient care and ultimately saving lives.
- **In transportation,** Convoy disrupted the trucking industry by introducing a machine learning-powered model to automate logistics and provide better matches for shippers and truckers, which drove efficiencies and lowered costs for both parties.
- **In the tech sector**, Intuit, creator of TurboTax, leverages computer vision and machine learning to help users file their taxes more efficiently. Using Amazon Textract, the solution lets users scan pictures of their tax forms, and the service verifies accuracy to identify any missing data or anomalies using contextual data from an existing database of tax codes and compliance forms.
- **In food services,** Domino's Pizza Enterprises Limited, a Domino's franchise holder with brands in Australia and Europe, created a predictive ordering solution to help stores anticipate what pizza their customers would order. The capability, part of a broader initiative to reduce pickup and delivery times, has made a difference: Pizzas are now ready for pickup in about three minutes or delivered within 10 minutes.

INNOVATION SPOTLIGHT

T-Mobile

At T-Mobile, engineering teams are using AI/ML technologies to transform customer support. Models leverage natural language processing to extract insights and relationships from unstructured text, as well as capabilities such as voice-to-text, text-to-speech, and language translation, with the goal of improving the customer experience.

Machine learning models sit between a care agent and the customer, scanning previous customer interactions and serving up relevant information to the care agent to help them quickly address the customer's issue. "We want to help those experts sitting in the call center solve your problem as quickly as possible using all the information that our company has at its disposal, in a way that is way more scalable than one human brain," says Heather Nolis, senior software engineer at T-Mobile.

Interestingly, success metrics for the technology are the same as for human agents, such as first-call resolution.

"We can measure based on just the accuracy of the models, but that doesn't actually help the business," says Nolis. "The things we are held accountable for are the same KPIs our care agents are held accountable for: how satisfied was the customer at the end of the call? Did they have to call back again to solve their problem? Because [improving] the way that our customers feel about our business and the way they experience our business is what we should be aiming for all the time."

Keys to scaling machine learning

Machine learning pilots are plentiful, but the challenge for many organizations is full-scale deployment. Here are five steps enterprises should consider to quickly and easily build, train, and deploy machine learning at scale:

Enlist executive buy-in from the business from the start. Too many proof-of-concept (POC) projects fail because they are started by small data science teams, who don't bring the business or executive teams in until after the system has been designed, instead of soliciting input at the onset of the project. A better process is to "work backwards," where the teams agree on a mutual set of business outcomes and design the solution and approach to achieve them.

"This approach provides a clear perspective on how you can go from Point A to Point B, helps you prioritize use cases, and what you should think about near-term versus recommendations for the long run," says Elaprolu.

Develop (and augment) your team. Highly specialized data scientists are critical, but you also need analysts who can translate business problems into ML solutions. With talent hard to find, companies need to balance recruitment with upskilling existing employees. T-Mobile, for example, has had success finding talent in unexpected places—including a marketing manager in the risk department who writes R code, and a Java developer with a master's degree in linguistics who helped lead a speech transcription project.

"Working very closely with the people who are going to consume what you make is the only way you can build something that's good and makes sense," says Nolis.

It's also important to augment in-house expertise with partners. This extends the community of experts you can tap into, while simultaneously offloading the "undifferentiated heavy lifting," so in-house engineers can focus on higher-value work.

> *"Working very closely with the people who are going to consume what you make is the only way you can build something that's good and makes sense."*
>
> ***— Heather Nolis, senior software engineer at T-Mobile***

Make sure you have enough data. The right data—and plenty of it—is critical to success when you move beyond a POC, which by definition focuses on testing a concept with a smaller set of data.

"In the real world, when you take a limited amount of data and try to deploy it to production, you will run into use cases and data sets that are not necessarily similar to what you've dealt with in the POC," says Elaprolu.

Create a process for continuous iteration. Conditions and business environments frequently change in the real world, so machine learning can't be a "set it and forget it" exercise.

"You need to have that continuous loop of retraining and redeploying ML models," says Elaprolu.

Offload undifferentiated heavy lifting. Instead of expending development resources to build capabilities from scratch, partner with a cloud services provider that can deliver core ML capabilities so your internal team can focus on higher-value activities.

For example, AWS AI helps developers augment applications with capabilities such as image and video analysis, natural language processing, personalized recommendations, and virtual assistants—with no machine learning experience required. In addition, tools like Amazon SageMaker simplify the process of building, training, and deploying models—and lower the barrier to entry. Learning tools like AWS DeepRacer, AWS DeepLens, and AWS DeepComposer can help developers learn ML fundamentals. AWS programs including the Machine Learning Solutions Lab and AWS Machine Learning Embark deliver hands-on onboarding, training, and implementation support to jump-start the process.

INNOVATION SPOTLIGHT

Formula 1

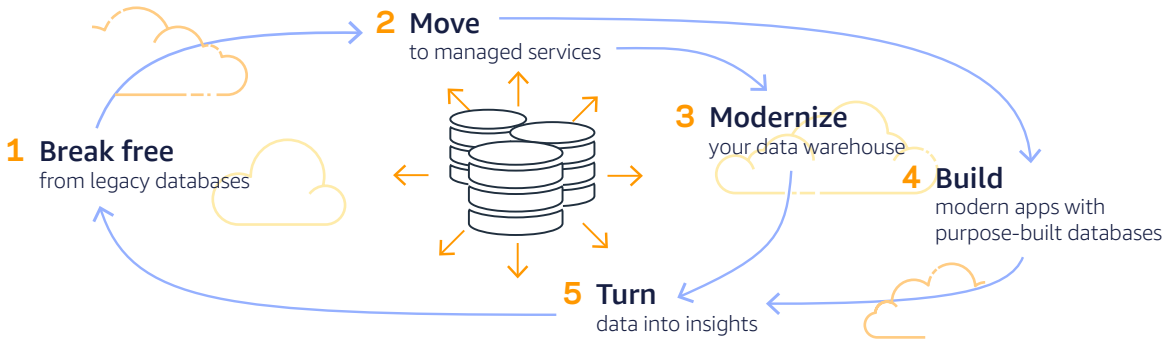
At Formula 1, machine learning is front and center in improving the experience for racing fans. Formula 1's engineering teams have built machine learning models that help fans better understand the split-second decisions made by a driver or pit crew that can dramatically affect the outcome of a race.

Machine learning dates back several years at Formula 1, where it was initially used to help racing teams better understand car performance, according to Rob Smedley, Director of Data Systems, Formula 1, and AWS technical ambassador. Smedley's team has since expanded its focus to fan experience, with many innovative applications that leverage analytics and ML, including F1 Insights powered by AWS and the F1 ProAm DeepRacer event.

"There's a lot of complexity and massive physics going on in the background," he says. "But the stuff that pops out on the screen is simple and engaging and brings about that excitement and that tangible entry point into all of these different stories that are going on as a race is unfolding."

While the use of analytics in Formula 1 is unique, the approach that Smedley's team takes with data and machine learning is applicable across industries.

"I think in any business, your first port of call is to look at what your problems are," Smedley says. "And once you understand your problems, then you put in place a strategy for fixing those problems."



5 steps to building a modern data foundation

Without an organizing framework, it's easy to be overwhelmed by the opportunities and challenges of building a modern, cloud-based data foundation. To help, AWS created the Data Flywheel, a holistic framework that applies the self-reinforcing loop principles set forth by Jim Collins to a data management strategy designed to maximize the value of data.

The Data Flywheel outlines five fundamental steps to building a modern, cloud-based data foundation:

1. **Break free from legacy databases**. Many organizations still have legacy, proprietary databases that are expensive, create lock-in, and carry punitive licensing terms. Moving to open-source databases can deliver cost efficiencies without causing a performance or availability hit.

2. **Move to managed services in the cloud**. As database platforms begin to scale up, IT time and administrative costs can grow as well. Cloud-based, managed database services reduce time spent on undifferentiated heavy lifting so teams can focus on higher-value activities.

3. **Modernize your data warehouse.** Traditional data warehouses don't have the ability to effectively store and analyze the growing volume and variety of data, which leads to data being stored in multiple silos. A modern "lake house" approach, including a data lake that can store unlimited volumes of data in various structured and unstructured formats, makes it much easier to catalog data, make it accessible, and analyze it across the business.

4. **Build modern apps with purpose-built databases.** Move from antiquated monolithic apps that run on one-size-fits-all relational databases to highly distributed microservice-based systems running on multiple purpose-built databases to solve each problem. This method frees the application from having to employ a single, overburdened database for every use case.

5. **Turn data into insights.** Data lakes, analytics, and machine learning help organizations gather smart, accurate insights faster and empower end users to see and visualize their data from any device or application.

The five steps are not linear, which gives organizations flexibility depending on their current level of data proficiency. "You can start anywhere, and they build on each other," says Oberoi.

Summary

Data, analytics, and machine learning have the potential to radically transform business processes and revenue models as well as shape future innovations. But data is only valuable if you can turn it into action.

To become a true data-driven organization, leadership teams need to shift culture to view data as a strategic asset. A modern, cloud-based infrastructure provides the scale, flexibility, and intelligence to support this shift and empower your business.

For more information, click here