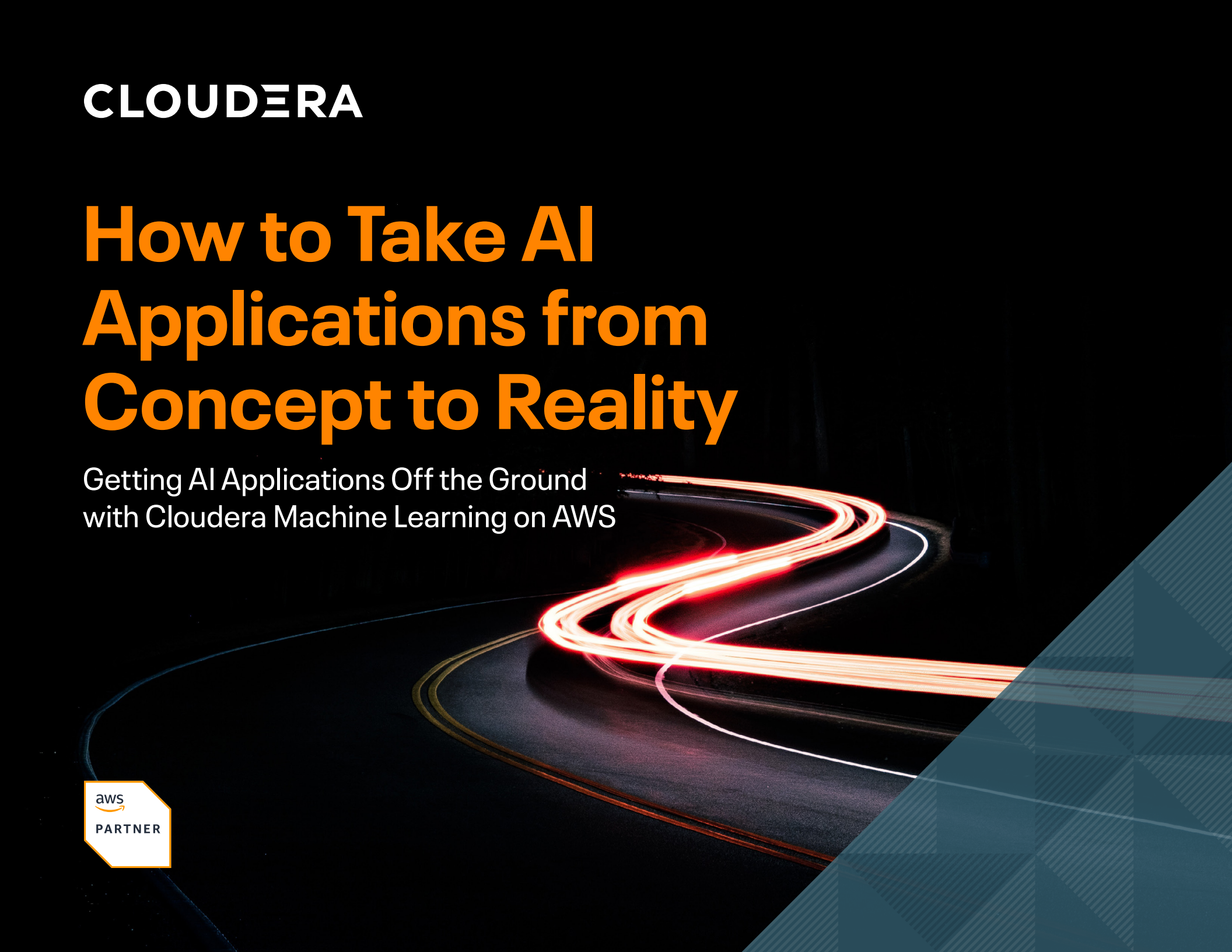
CLOUDERA

How to Take AI Applications from Concept to Reality

Getting AI Applications Off the Ground with Cloudera Machine Learning on AWS



aws PARTNER

Table of Contents

Introduction: Preparing for AI with an AWS Cloud Strategy

Whatever industry you're in, business challenges are arriving faster than ever. Increasingly it takes Artificial Intelligence and Machine Learning tools to keep pace. AI/ML applications, which can consume intense compute and data resources, are a natural fit for the public cloud services of Amazon Web Services (AWS).

Today AI/ML applications in AWS are widely used to:

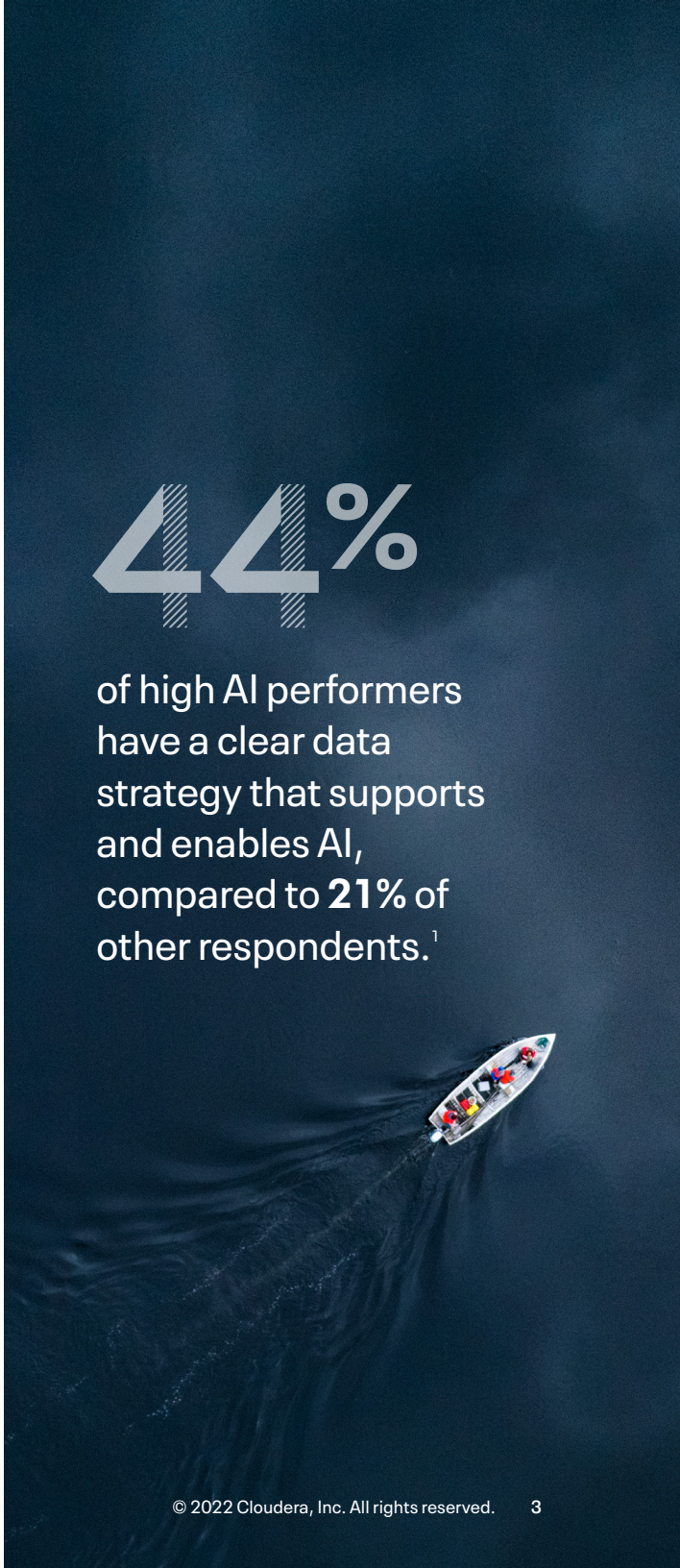
- prevent fraud
- analyze patient data
- improve customer service
- boost manufacturing efficiency
- extend the life of equipment
- optimize supply chains

And that's just scratching the surface. In any field that relies on data, AI and ML can help solve difficult use cases efficiently, finding patterns in data sets and using them to make predictions—saving time and helping businesses stay competitive. These competitive advantages come from putting AI/ML into practice quickly with cloud resources, a key benefit of AWS.

Your organization may feel pressure to deliver a public cloud strategy that puts AI applications into practice—not just in cutting-edge demonstrations that never leave the lab, but in production. A unified data platform on AWS can help solve that critical step of moving AI from the lab to the factory. Without it, your teams may adopt point solutions that don't work together and lead to greater organizational problems. An integrated AI/ML lifecycle, running on a modern enterprise data platform in an AWS environment, makes it possible to put powerful technologies to use right away, at the moments when they'll make the greatest impact.

In this eBook we'll examine the challenges of getting AI applications off the ground, and ways to overcome those challenges using a data platform built on AWS's public cloud resources.

1 McKinsey, "The state of AI in 2020," 17 November 2020



Chapter 1:
Data is Your Foundation

It's a given that AI and ML require ready access to huge volumes of data. How you navigate the data lifecycle to deliver public cloud applications that make business-changing predictions is where complexity arises.

Your ideal lifecycle includes training models with big, validated, current datasets. From there, your ML applications are applied to real-time data streams, to make predictions in the moment. And, as we'll detail later, all of this happens without compromising on security and governance.

These voracious demands for ever-more data, and the need for scalability, mean the enterprise data landscape is increasingly adopting cloud environments. Public cloud providers play a significant role in supporting ML strategies.

Cloudera Data Platform (CDP) with Cloudera Machine Learning (CML) is designed for the level of scalability needed today. CML delivers a full toolset for ML, forming a foundation for AI applications that can empower all users, technical or nontechnical, to access trusted data and use common tools to accelerate data-driven insights.

Users can experience CML's full potential cost effectively through AWS. CML on AWS offers customers a managed and supported suite of products and services, complete with security and governance.

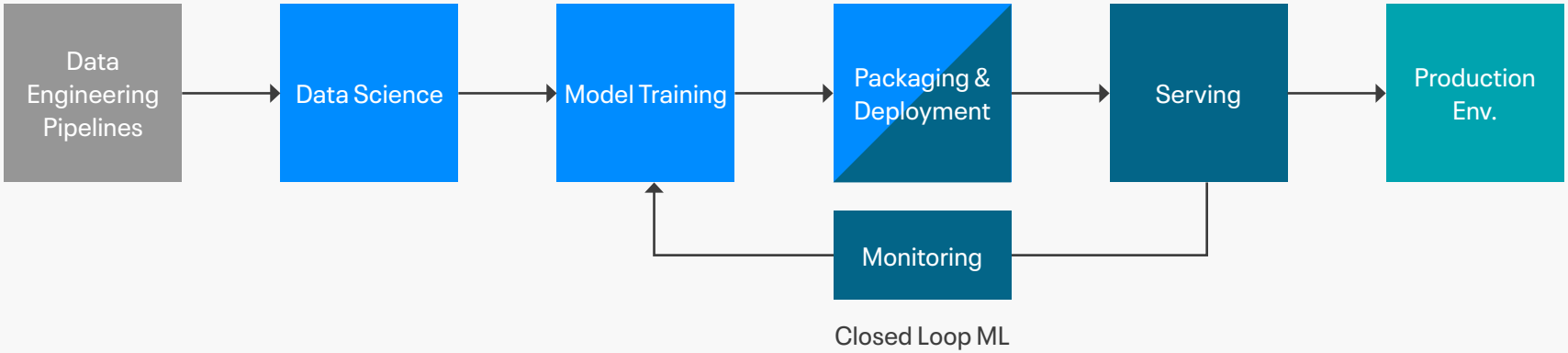
CML on AWS Success Story: Globe Telecom

Globe Telecom, a leading mobile and broadband provider in the Philippines, runs machine learning analytics using infrastructure built on CDP running on AWS. Taking advantage of this modern analytics environment, Globe Telecom is able to perform customer segmentation to enhance customers' mobile experiences and deliver relevant advertising. Globe Telecom manages 600PB of data with CML on AWS, driving a strategy that supports 66% year-over-year growth in mobile data volume.

Read more here

Machine Learning at work

An integrated lifecycle supports ML in production.



Closed Loop ML

Chapter 2: Start Fast with Pre-Built Machine Learning Projects

Plenty of organizations that understand the value of AI and ML still struggle to get applications into production quickly and at scale. Giving data teams access to a wealth of data on a public cloud platform is just one part of the process. Workflow struggles mean useful ML models can take weeks to deploy, or never make it to production at all. Just 35% of organizations say analytical models are fully deployed in production.[2]

Pre-built prototypes and open-code business applications can help reduce development snags that too often prevent AI applications from being useful.

40%

of companies take more than 30 days to deploy a single ML model.[3]

2 MESA, "IDC: AI Can Significantly Help Organizations with Analytics, Business Intelligence," 27 September 2019

3 Algorithmia, "The 2020 state of enterprise machine learning,"

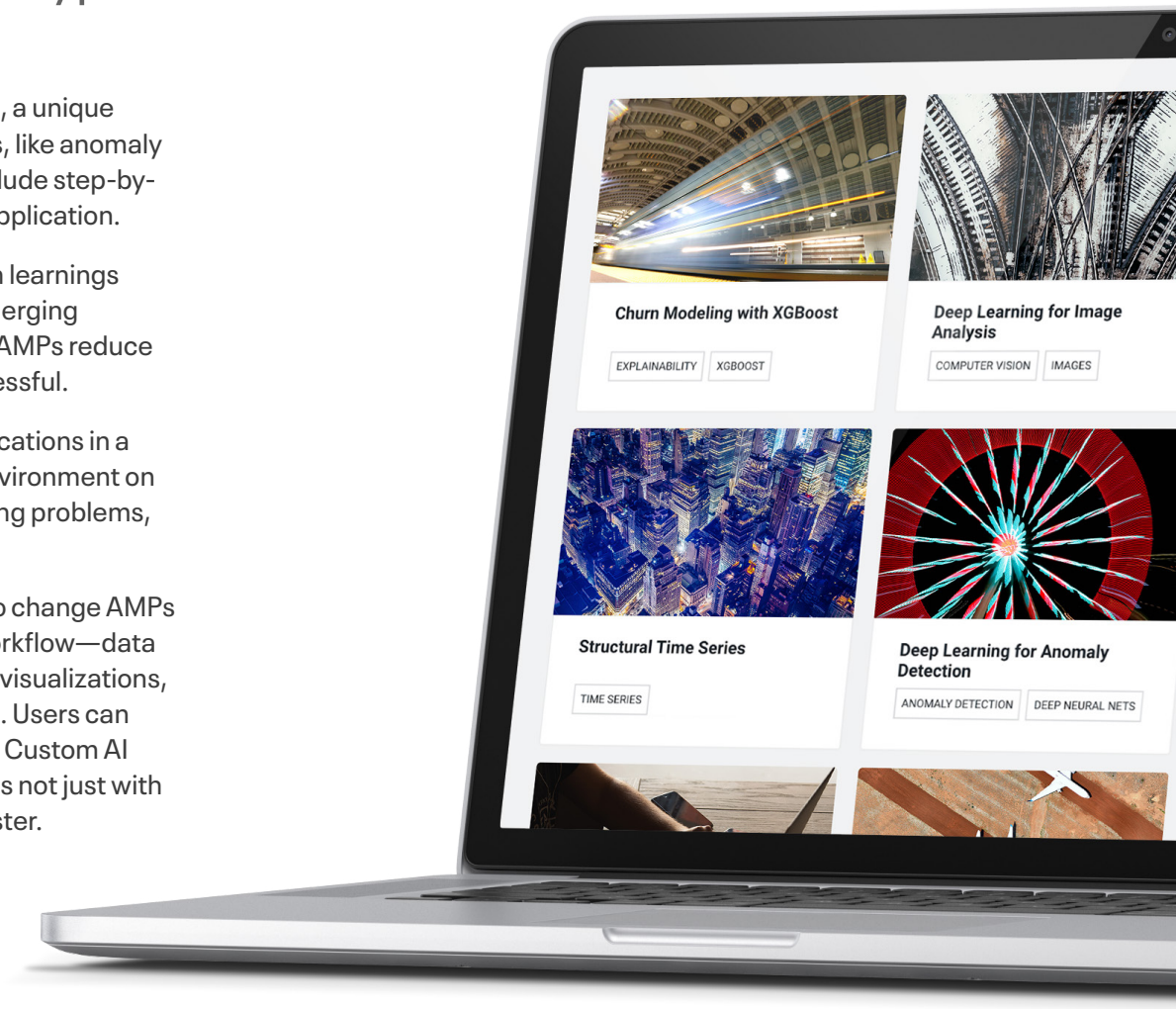
Change the way ML projects are built and delivered with fully-developed prototypes

CML on AWS offers Applied Machine Learning Prototypes (AMPs), a unique collection of pre-built models around common industry use cases, like anomaly detection, churn modeling, and visual object detection. AMPs include step-by-step guidance and end-to-end workflows from data to model to application.

Cloudera data scientists and researchers develop AMPs based on learnings from Cloudera Fast Forward Labs Research, which focuses on emerging trends across the data science and machine learning landscape. AMPs reduce development snags that prevent AI applications from being successful.

CML on AWS users can use AMPs to deploy fully working ML applications in a single click. Examples can be automatically deployed in a CML environment on AWS. AMPs incorporate best practices for solving machine learning problems, with steps defined in YAML configuration files.

Pre-built does not mean one-size-fits-all. Users are empowered to change AMPs to suit their business needs, and can customize their entire ML workflow—data ingestion, feature engineering, model training, model publishing, visualizations, and building interactive web applications to communicate results. Users can even borrow code and use it in a completely different application. Custom AI applications are ready to deploy in a fraction of the time. This helps not just with scale, but with speed—serving to deliver value to the business faster.



Chapter 3: Bring Teams Together with Better Collaboration

Even with scalability and speed, AI applications won't deliver value if they aren't easily accessible to people who can use them. Deploying AI through the organization means ensuring the right people have access.

For ML and data engineers, the focus is on production and operation workflows. They need an integrated experience that supports orchestrated, automated data pipelines, making sure teams have the data workflows they need to collaborate on ML projects.

Data scientists need access to a versatile and complete development environment. Much of this work needs to take place in isolated and containerized workspaces, while giving data scientists the flexibility to run any IDEs, libraries, or frameworks they choose.

And business users need access to analytics they can use to make confident data-driven decisions. They benefit from secure visual applications that display results backed by full auditability.

CML on AWS includes tools to help you to get to production and scale your AI use cases in the most effective way possible. ML and data engineers, data scientists, and business users can all access the data, tools, and environments they need to deliver applications in the most effective ways. This includes built-in APIs that developers can use to incorporate ML predictions into other applications. And CML on AWS's preconfigured runtimes offer flexible, containerized environments that developers can use to access ML resources and get customized AI applications running quickly.

CML Success Story: Western Union

Processing 29 transactions per second, Western Union's data set exceeds 100 terabytes. With an enterprise data hub from Cloudera, Western Union is able to support 60X faster data loading, enabling predictive analytics on structured and unstructured data sets at the time of the transaction.

Read more here

Chapter 4: Turn Data into Action with Data Visualization

ML insights can go overlooked if they're not clearly understandable and readily available—the challenges that visualizations can solve. Data visualizations are instrumental in eliminating knowledge gaps and bridging connections between stakeholders. And modern data visualization extends beyond just dashboards to include automated reporting and predictive applications.

Data visualization supports:

Share insights everywhere. The ability to set up and share dashboards easily means data gets to the right people faster. With a data visualization solution, users can build and publish custom dashboards with easy-to-use web-based tools.

Automate intelligent reporting. Reporting, whether on a regular cadence or in response to an event, can increase data consumption. Everyone can have access to the latest insights with scheduled updates, emailed reports, and dynamic alerts.

Build predictive applications. The value of predictive analytics grows when business users can access the predictions directly. Visualizations built on machine learning models can enable everyone to ask predictive questions and get real-time insights.

66%

of IT executives view AI as critical to success.[4]

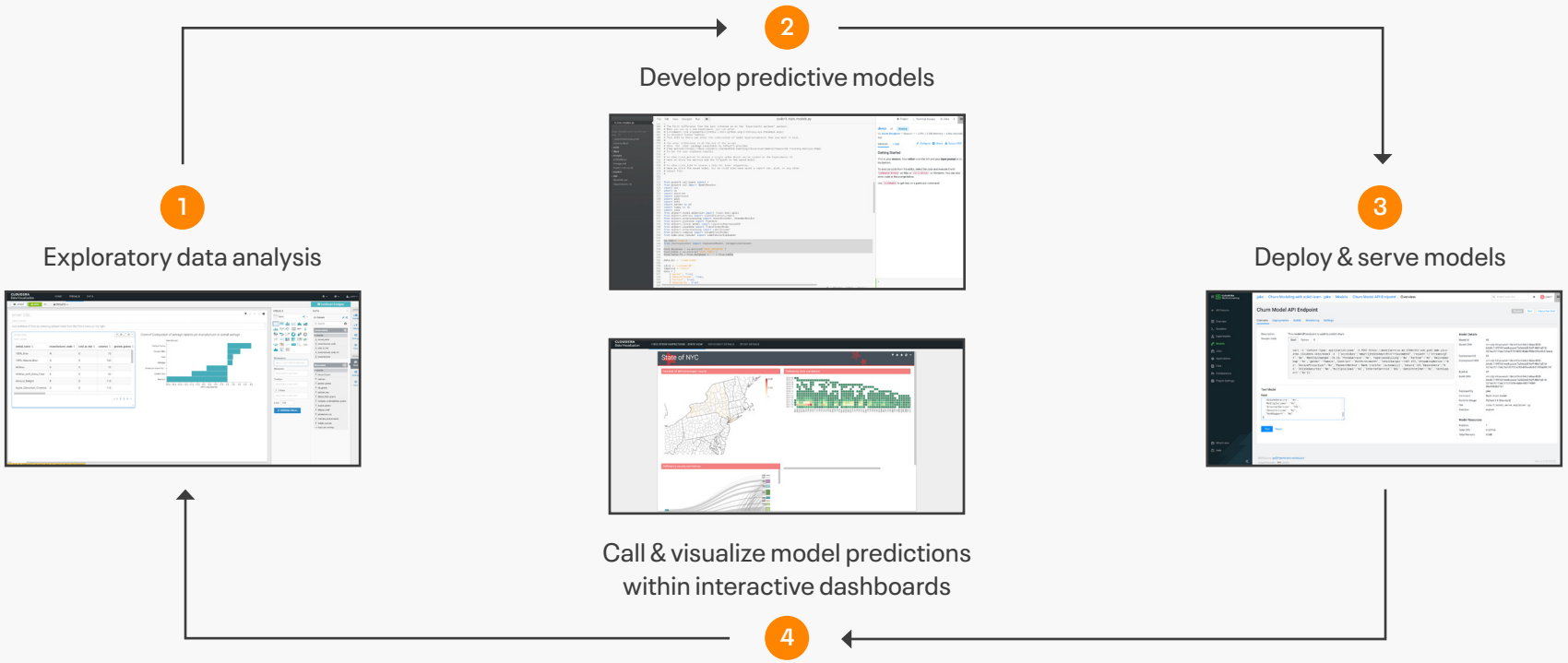
4 Deloitte, "State of AI in the Enterprise, 4th Edition," 2020

Cloudera Data Visualization (CDV) lets AI and ML stakeholders create visual objects to explore data and communicate insights. With speed and time-to-value in mind, visualizations help streamline model publishing and get ML analytics into usage fast.

CDV offers self-service data visualization workflows, with a web-based, no-code, drag-and-drop user interface. For deeper insights, data practitioners can build advanced dashboards, serving sophisticated insights at a glance.

Users can share the same visualization tools across the platform, connecting the dots between raw data, production ML workflows, and business impact.

Visualizing ML



A managed and supported platform running on AWS

Chapter 5: Manage the Risks of Security and Governance

Data breaches are increasingly common, and privilege abuse and data mishandling are the most common kinds of misuse that lead to a breach.[5] For anyone working with AI and ML, data security and governance should be a primary consideration, not a nice-to-have. But as enterprises begin exploring the value of AI applications, users who decide to deploy their own solutions for data experimentation can introduce new risks.

Cloudera makes security and compliance an integrated part of the data lifecycle. Shared Data Experience (SDX), a core part of Cloudera Data Platform, provides an integrated set of security and governance technologies, run independently from compute and storage layers. SDX provides a robust set of tools to deliver consistent data context across deployments, through automatic model cataloging and lineage, along with governed and secure production workflows. Data lineage, management, and automation are built in.

Security risks and data privacy requirements are important from the start, since even experimental AI applications could one day end up in production. The risks loom large, with the potential of costly data breaches, reputational damage, and fines from regulators if things go wrong.

CML on AWS includes support for security and governance, empowering ML practitioners to build and share ML models that drive value, while keeping data in place, to reduce additional risks to security or compliance.

Cloudera Data Platform's built-in, always-on SDX layer gives you full control and visibility into:

- Security
- Governance
- Lineage
- Management
- Automation

80%

of organizations seeking to scale digital business will fail because they do not take a modern approach to data and analytics governance.[6]

5 Verizon, "2021 Data Breach Investigations Report"

6 Gartner, "Our Top Data and Analytics Predicts for 2021," 12 January 2021

Conclusion: Simplicity or Complexity? Find Your Sweet Spot.

For some businesses, the roadmap to AI success includes sophisticated, custom app development, with ML models trained on billions of data points. For others, getting working AI applications into production quickly is what matters most, and fast access to data, preconfigured runtimes, and step-by-step guidance can make all the difference.

Whether you choose pre-built components or develop your own, it takes the right data foundation to deploy enterprise AI applications that make an impact. CML on AWS enables you to build the ML lifecycle that's right for your needs in one complete solution.

Point solutions may solve some AI challenges, but they can also create functional silos, cause vendor-lock-in, and introduce other such inefficiencies. CML on AWS enables a complete ML lifecycle in an open, hybrid cloud architecture, with easily accessible tools.

It's all designed to help data science teams and business leaders improve collaboration, deliver more models faster, and drive immediate business actions.

Take Your Next Step

Cloudera Data Platform Machine Learning on Amazon Web Services offers a fast, cost-effective way to stand up machine learning applications in the public cloud. CML on AWS is a full suite of products and services on a single, integrated modern data platform.

Read more

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com | US: +1 888 789 1488 | Outside the US: +1 650 362 0488

Sources

1 McKinsey, "The state of AI in 2020," 17 November 2020

2 MESA, "IDC: AI Can Significantly Help Organizations with Analytics, Business Intelligence," 27 September 2019

3 Algorithmia, "The 2020 state of enterprise machine learning,"

4 Deloitte, "State of AI in the Enterprise, 4th Edition," 2020

5 Verizon, "2021 Data Breach Investigations Report"

6 Gartner, "Our Top Data and Analytics Predicts for 2021," 12 January 2021

CLOUDERA