Unlocking the value of your data at scale in the financial services industry

AWS Prescriptive Guidance

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS Prescriptive Guidance: Unlocking the value of your data at scale in the financial services industry

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Unlocking the value of your data at scale in the financial services industry

Brian Cavanagh, Junaid Baba, Sarah Bourial, Sokratis Kartakis, Dhrubajyoti Mukherjee, Maren Suilmann, Maira Ladeira Tanke, Pauline Ting, and Amine Ait el harraj, Amazon Web Services

The financial services (FS) industry is facing major disruption from financial technology companies (fintechs) and digital-only banks that are leading the way in the digital transformation of banking. This transformation is increasingly characterized by the development of banking products and services that are based on artificial intelligence and machine learning (AI/ML) technologies. According to AI-bank of the future: Can banks meet the AI challenge? from McKinsey & Company, FS organizations must deploy AI/ML technologies at scale if they want to stay relevant. To deploy AI/ML technologies at scale, FS organizations must first unlock the business value of their data.

FS organizations hold vast amounts of data but struggle to extract value from this data at scale. By unlocking the value of data, FS organizations can deliver deeper and more personalized offerings, insights, and support to their customers and partners. The unlocked value can also help FS organizations quickly expose inefficiencies in current processes from capital markets to manufacturing operations, while providing prioritized insights into areas that require optimization. This strategy describes how you can unlock the value of your data at scale by developing ML capabilities in your organization. The intended audience for this strategy includes CEOs, CFOs, CIOs, and senior managers in the banking and wealth management industry.

This strategy helps you understand the following:

- Business outcomes for launching ML capabilities in your organization
- Metrics and target scores for operational success
- A scalable ML framework for transforming your organization's ML capabilities
- AWS best practices for scaling (based on hundreds of customer implementations)

Targeted business outcomes

The key business outcome is to adopt optimized processes that can transform your organization's data into business value. Specifically, this strategy can help you achieve the following targeted business outcomes:

- Scale machine learning and operations (MLOps) capabilities across your organization and develop an ML platform to support hundreds of data scientists and data engineers to run ML workflows in your organization.
- Implement a scalable, secure, cost-efficient, and sustainable infrastructure deployment by using AWS Service Catalog to create an on-demand managed infrastructure with Amazon SageMaker.
- Standardize the ML model development and deployment process across multiple teams.
- Reduce the technical debt of existing models and create reusable artifacts to speed up future model development.
- Reduce idea-to-value time for data and analytics use cases down to 12–16 weeks.
- Reduce the creation time for ML use case environments down to 1–2 days.

When you scale the use of advanced analytics in the enterprise, it can take too long to develop and release ML models and solutions into production. By partnering with AWS, you receive guidance and support that can help you rapidly design, build, and launch a modern, secure, scalable, and sustainable self-service platform for developing and productionizing ML-based services to support your business and customers. AWS Professional Services can work with you to accelerate the implementation of AWS best practices for using Amazon SageMaker AI to build, train, and deploy ML models for use cases customized to your organizational needs.

The collaboration provides the following:

- A federated, self-service, and DevOps-driven approach for infrastructure and application code, with a clear route to live that can reduce deployment times from weeks to minutes
- A secure, controlled, and templated environment to accelerate innovation with ML models and insights that use industry best practices and bank-wide shared artifacts
- The ability to access and share data more easily and consistently across your organization
- A modern toolset based on an on-demand managed architecture that can minimize compute requirements, drive down costs, and enable sustainable ML development and operations, with

the flexibility to accommodate new AWS products and services to meet ongoing use case and compliance requirements

- Adoption, engagement, and training support that enables data science and engineering teams across your organization to be self-sufficient and scale ML products across your organization

Operational metrics

AWS Professional Services can work with your internal teams to establish a delivery governance mechanism for tracking your organization's progress against the operational metrics defined in the following table. The metrics are based on real-world values and can help define your MLOps engagement. The values in the following table are guidelines only. The target values will depend on your organization's implementation.

Area	Sub area	Metric	6 month target from platform launch
1. Operational efficiency	Faster delivery	Time to value	3 month average
	Cost transparency through clear cost allocations to business divisions	AWS management costs	Less than initial
2. Boundary-less delivery	Freedom of access seamlessly granted in a controlled manner	Time to access	1 day
	Support for CI/CD	Time to value (idea to live)	3 month average
	Simplified route-to-live delivery	Time to deploy (post beta)	3 month average
	On-demand lab spin up/down	Time to spin up environment	24 hours
3. Controlled	Access aligned to each use case and granted or removed dynamically	Time to access use case	< 1 day
	Persistent production environment	Time to deploy	2 weeks

	allowing spoke-driven development to live		
	Storage and compute costs managed at point of use and attributed to cost center	Actuals/forecast transparent	Forecast accuracy variance permitted
	Compliance with internal security standards (automatic assessment and auditing)	Number of non-compliant environments	0
	Data owners accountable for contributing and sharing data (within SLA)	Copies of data in data lake	0 copies of data in data lake
	Common processing with support for regulatory separation of data	Compliant access	0 regulatory findings
4. Constantly evolving and repeatable patterns	A model development environment with constantly evolving standards	Number of updates per year	6
	Ability to package and deploy environments at will by using spokes	Time to spin up environment	2 hours

	Clear layout of services and infrastructure being consumed	Non-tracked infrastructure	0
5. For all of the bank	Reduced number of distinct modelling environments	Number of environments	< 2
	Simplified team collaboration across organizational boundaries	Time to access (for example, data)	1 day
	Common and singular operating model/framework	Number of domains controlled by model	> 3

Solution overview

A scalable ML framework

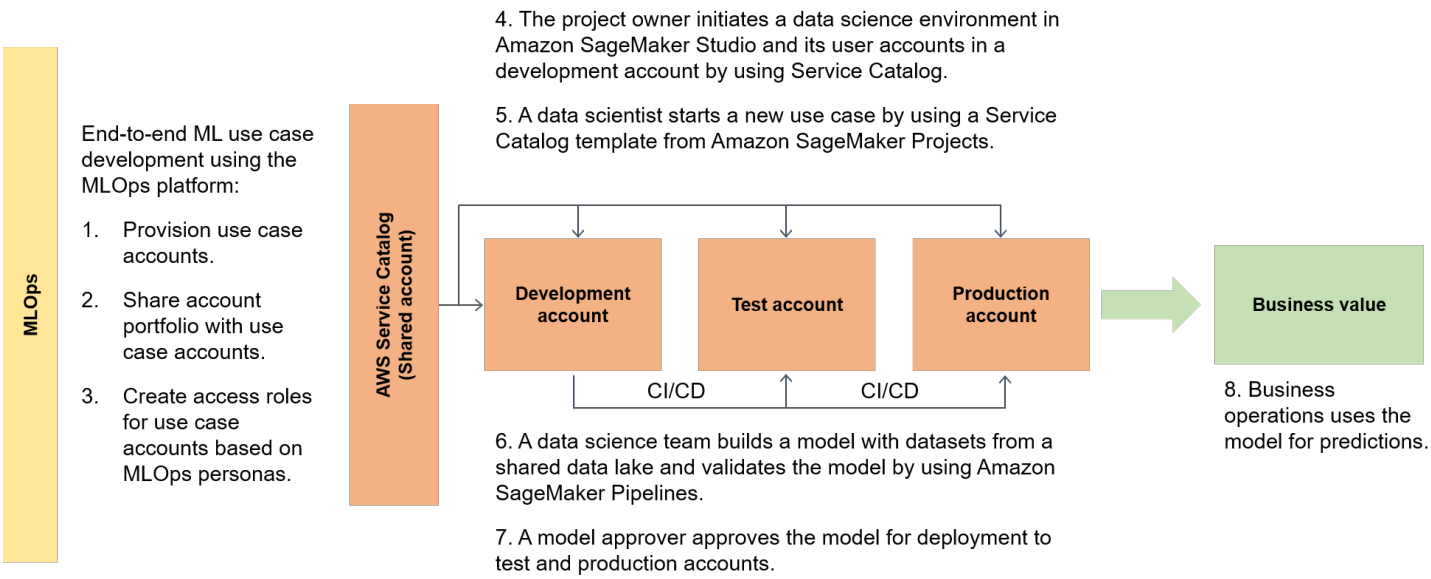
In a business with millions of customers spread across multiple lines of business, ML workflows require the integration of data owned and managed by siloed teams using different tools to unlock business value. Banks are committed to the protection of their customers' data. Similarly, the infrastructure used for the development of ML models is also subject to high security standards. This additional security adds further complexity and impacts the time to value for new ML models. In a scalable ML framework, you can use a modernized, standardized toolset to reduce the effort required for combining different tools and simplify the route-to-live process for new ML models.

Traditionally, the management and support of data science activities in the FS industry is controlled by a central platform team that gathers requirements, provisions resources, and maintains infrastructure for data teams across the organization. To rapidly scale the use of ML in federated teams across the organization, you can use a scalable ML framework to provide self-service capabilities for developers of new models and pipelines. This enables these developers to deploy modern, pre-approved, standardized, and secure infrastructure. Ultimately, these self-service capabilities reduce your organization's dependency on centralized platform teams and speed up the time to value for ML model development.

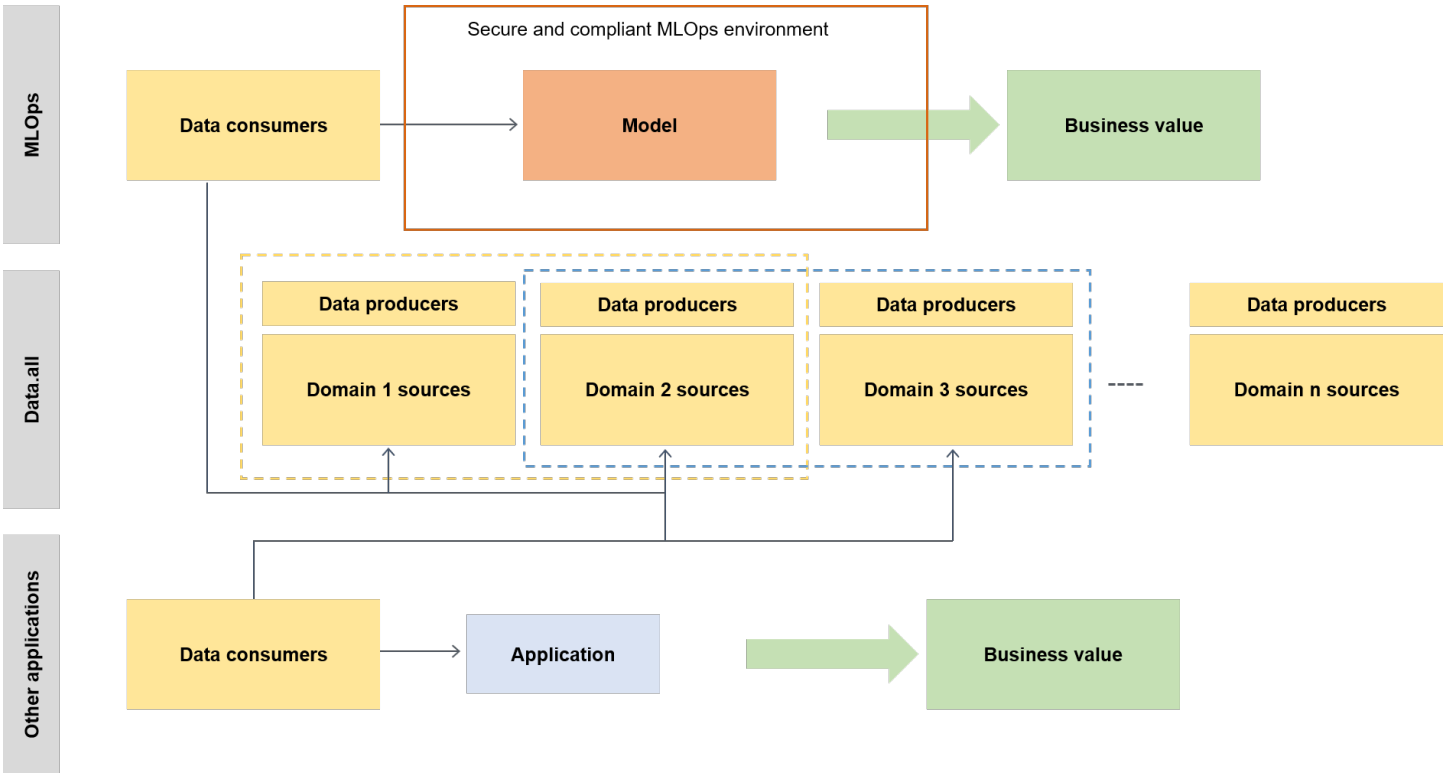
The scalable ML framework enables data consumers (for example, data scientists or ML engineers) to unlock business value by giving them the ability to do the following:

- Browse and discover pre-approved data that's required for model training
- Gain access to pre-approved data quickly and easily
- Use pre-approved data to prove model viability
- Release the proven model to production for others to use

The following diagram highlights the end-to-end flow of the framework and the simplified route to live for ML use cases.



In the wider context, data consumers use a serverless accelerator called *data.all* to source data across multiple data lakes and then use the data to train their models, as the following diagram illustrates.



At a lower level, the scalable ML framework contains the following:

- **Self-service infrastructure deployment** – Reduce your dependency on centralized teams.

- **Central Python package management system** – Make pre-approved Python packages available for model development.
- **CI/CD pipelines for model development and promotion** – Decrease time to live by including continuous integration and continuous (CI/CD) pipelines as part of your infrastructure as code (IaC) templates.
- **Model testing capabilities** – Take advantage of unit testing, model testing, integration testing, and end-to-end testing functionalities that are automatically available for new models.
- **Model decoupling and orchestration** – Avoid unnecessary compute and make your deployments more robust by decoupling model steps according to computational resource requirements and orchestration of the different steps by using Amazon SageMaker Pipelines.
- **Code standardization** – Improve the quality of your code by using CI/CD pipeline integration for validating Python Enhancement Proposal (PEP 8) standards.
- **Quick-start generic ML templates** – Get Service Catalog templates that instantiate your ML modelling environments (development, pre-production, and production) and associated pipelines with the click of a button by using SageMaker Projects for deployment.
- **Data and model quality monitoring** – Make sure that your models perform to operational requirements and within your risk-tolerance level by using Amazon SageMaker Model Monitor to automatically monitor drift in your data and model quality.
- **Bias monitoring** – Enable your model owners to make fair and equitable decisions by automatically checking for data imbalances and if changes in the world have introduced bias to your model.

A central hub for metadata

Data.all is a serverless accelerator that you can integrate with your existing AWS data lakes to gather metadata into a central hub. A simple, easy-to-use UI in data.all displays metadata associated with datasets from multiple, existing data lakes. This enables both non-technical and technical users to search for, browse through, and request access to valuable data that can be used in their ML labs. Data.all uses AWS Lake Formation, AWS Lambda, Amazon Elastic Container Service (Amazon ECS), AWS Fargate, Amazon OpenSearch Service, and AWS Glue.

SageMaker validation

To prove the capabilities of SageMaker across a range of data processing and ML architectures, the team implementing the capabilities selects, together with the banking leadership team, use cases of varying complexity from different divisions of banking customers. The use case data is obfuscated and made available in a local Amazon Simple Storage Service (Amazon S3) data bucket in the use case development account for the capabilities proving phase.

When the model migration from the original training environment to a SageMaker architecture is complete, your cloud-hosted data lake makes the data available to be read by the production models. The predictions generated by the production models are then written back to the data lake.

After the candidate use cases have been migrated, the scalable ML framework takes an initial baseline for the target metrics. You can compare the baseline to previous on-premises or other cloud provider timings as evidence of the time improvements enabled by the scalable ML framework.

Best practices

This section provides an overview of AWS best practices for MLOps.

Account management and separation

AWS best practices for account management recommend that you divide your accounts into four accounts for each use case: experimentation, dev, test, and prod. It's also a best practice to have a governance account for providing shared MLOps resources across the organization and a data lake account for providing centralized data access. The rationale for this is to completely separate the development, test, and production environments, avoid delays caused by service limits being hit through multiple use cases and data science teams sharing the same set of accounts, and provide a complete overview of the costs for each use case. Finally, it's a best practice to separate account-level data, as each use case has its own set of accounts.

Security standards

To meet security requirements, it's a best practice to turn off public internet access and encrypt all data with custom keys. Then, you can deploy a secure instance of Amazon SageMaker Studio to the development account in a matter of minutes by using Service Catalog. You can also get auditing and model monitoring capabilities for each use case by using SageMaker through templates deployed with SageMaker Projects.

Use case capabilities

After the account setup is complete, your organization's data scientists can request a new use case template by using SageMaker Projects in SageMaker Studio. This process deploys the necessary infrastructure to have MLOps capabilities in the development account (with minimal support required from central teams), such as CI/CD pipelines, unit testing, model testing, and model monitoring.

Each use case is then developed (or refactored in the case of an existing application code base) to run in a SageMaker architecture by using SageMaker capabilities such as experiment tracking, model explainability, bias detection, and data/model quality monitoring. You can add these capabilities to each use case pipeline by using pipelines steps in SageMaker Pipelines.

MLOps maturity journey

The MLOps maturity journey defines the necessary MLOps capabilities made available in a company-wide setup to ensure that an end-to-end model workflow is in place. The maturity journey consists of four stages:

1. **Initial** – In this stage, you establish the experimentation account. You also secure a new AWS account within your organization where you can experiment with SageMaker Studio and other new AWS services.
2. **Repeatable** – In this stage, you standardize code repositories and the ML solution development. You also adopt a multi-account implementation approach, and standardize your code repositories to support model governance and model audits as you scale out the offering. It's a best practice to adopt a production-ready model development approach with standard solutions that are provided by a governance account. The data is stored in a data lake account and use cases are developed in two accounts. The first account is for experimentation during the data science exploration period. In this account, data scientists discover models for solving the business problem and experiment with multiple possibilities. The other account is for development, which takes place after the best model has been identified and the data science team is ready to work in the inference pipeline.
3. **Reliable** – In this stage, you introduce testing, deployment, and multi-account deployment. You must understand MLOps requirements and introduce automated testing. Implement MLOps best practices to ensure models are both robust and secure. During this phase, introduce two new use case accounts: a test account for testing models developed in an environment that emulates the production environment and a production account for running model inference during business operations. Finally, use automated model testing, deployment, and monitoring in a multi-account setup to ensure that your models meet the high quality and performance bar that you set.
4. **Scalable** – In this stage, you templatize and productionize multiple ML solutions. Multiple teams and ML use cases start to adopt MLOps during the end-to-end model building process. To achieve scalability in this stage, you also increase the number of templates in your template library through contributions from a wider base of data scientists, reduce time to value from idea to production model for more teams across the organization, and iterate as you scale.

For more information on the MLOps maturity models, see MLOps foundation roadmap for enterprises with Amazon SageMaker on the AWS Machine Learning Blog.

A vision of success

To successfully unlock the value of your data, your organization must develop a scalable ML framework that achieves your business outcomes while aligning to the AWS best practices defined in this strategy. Such a framework could help your organization achieve the following:

- A self-service process for creating standardized production-level infrastructure for developing advanced analytics, data science-based workloads, and a clear DevOps-driven route to live for those workloads
- Use case teams that can enjoy streamlined governance, which reduces time to value while also freeing up valuable technical resources
- A data science library to access and share reusable ML templates that can rapidly reduce end-to-end delivery timescales by using templates that can be refined and shared with others
- A cloud data reference library (DataHub) that provides metadata on all data located across multiple AWS data lakes, speeding up not only data discovery but time to access the data required to train models
- A dedicated DevOps team trained to manage the platform and underlying toolkit, support new feature development, and ensure ongoing compliance with key controls
- Real-life use cases delivered through to production-level accounts
- A knowledge transfer between the project team and the receiving business line teams (spoke teams), which enables business teams to self-manage continued development and the route-to-live process
- Completion of initial engagement and adoption activities with a targeted group of business teams
- Hundreds of classroom-based AWS training sessions delivered to support adoption

FAQ

When should I build an MLOps platform?

It's time to standardize on an MLOps platform when you notice that your engineers are spending more time on researching and seeking approvals for tooling options than they are on building ML models.

Can I integrate other ML tools into the MLOps platform?

Yes. You can integrate non-AWS tools into the platform. While SageMaker Studio is at the core of the MLOps platform, you can still integrate other products with the SageMaker Studio suite of services.

How can my organization simplify governance requirements to accelerate innovation?

As part of the use case candidates that you select to prove the build of your MLOps platform, ensure that the use cases are of sufficient complexity, require various data classifications, and require big data volumes. By doing this, not only do you prove the platform capability, but you do the heavy lifting from a governance perspective as part of your initial platform release. If you can do this, then teams that adopt the MLOps platform as part of the rollout will have a lighter governance load, as they use a platform that has already met the governance requirements for complex use cases.

Which team do I need in place for building an MLOps platform?

A robust MLOps foundation, which clearly defines the interaction among multiple personas and technologies, can increase time to value, reduce cost, and enable data scientists to focus on innovation. Having the right team can be the difference between failure and success for MLOps platform development. Due to the nature of MLOps, many roles need to be involved, such as data scientists, ML engineers, DevOps professionals, data owners, IT owners, business analysts, and product owners. Make sure that all your stakeholders are interacting in a cross-functional team to ensure the best outcomes for your MLOps platform.

How can I start on my MLOps journey?

You can start by creating a secure experimentation environment where data scientists receive a snapshot of data. The data scientists can use SageMaker to experiment and ultimately prove that ML can solve a specific business problem.

Should an MLOps transformation be driven by a top-down or bottom-up approach in an organization?

While bottom-up approaches can be successful, support from leadership is essential for the success of MLOps platform development. With a top-down approach, you can ensure faster standardization of the developed solution, reduce costs, and achieve higher scalability and reusability between the models developed by different teams in your organization.

Next steps

To unlock the business value of your data by adopting a scalable ML framework, consider the following next steps:

1. Contact an AWS sales representative or contact an AWS Professional Services representative.
2. Pursue training and certification opportunities for the appropriate team members in your organization:
   - AWS training and certifications
   - MLOps Engineering on AWS
   - AWS Certified Big Data - Specialty
   - AWS Certified Machine Learning - Specialty
   - AWS Certified DevOps Engineer - Professional
   - AWS Private Training - Information Request

Resources

- What's New with Machine Learning? (AWS documentation)
- Automate MLOps with SageMaker Projects (YouTube)
- Put your data to work on the most scalable, trusted, and secure cloud (AWS documentation)
- MLOps foundation roadmap for enterprises with Amazon SageMaker (AWS Machine Learning Blog)
- Part 1: How NatWest Group built a scalable, secure, and sustainable MLOps platform (AWS Machine Learning Blog)
- AI-bank of the future: Can banks meet the AI challenge? (McKinsey & Company)

Document history

The following table describes significant changes to this guide.

Change	Description	Date
Update	Replaced DataHub with data.all.	November 7, 2023
Initial publication	—	September 16, 2022