Introduction to MLOps

Bringing DevOps and Automation to Machine Learning

Hei Chow
Solutions Architect

Current state of AI/ML

State of machine learning

• Today

- 53% of POCs make it into production
- Average 9 months

- Gartner

Last decade

- Focusing mostly on building ML models
- Operationalization was an afterthought

By end of 2024

- 75% of organizations will shift from piloting to operationalizing AI

- Gartner

https://www.idgconnect.com/article/3583467/gartner-accelerating-ai-deployments-paths-of-least-resistance.html

Main Challenges

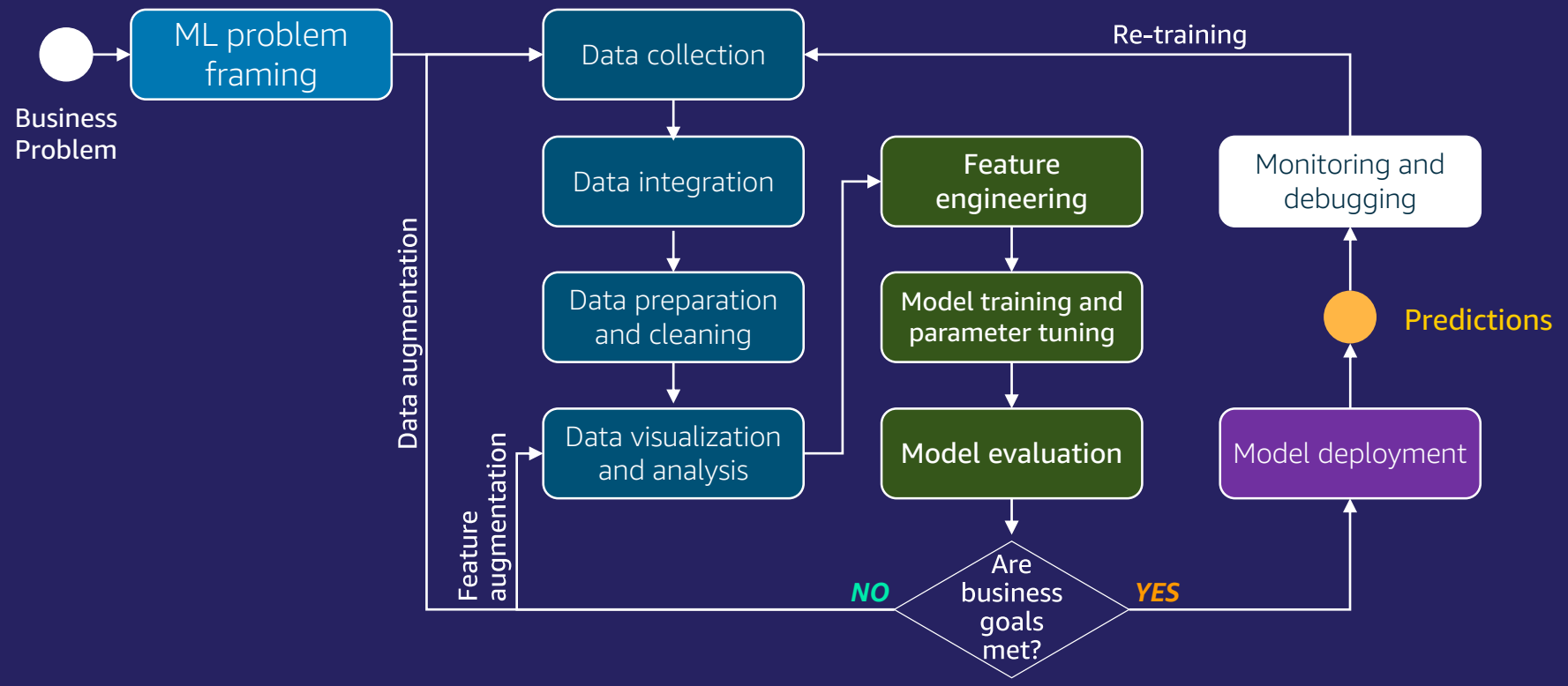
- Publishing a ML model is not enough.
- Managing the published ML models is as important as developing them.

- *"IT leaders responsible for AI are discovering **'AI pilot paradox'**, where launching pilots is deceptively easy but deploying them into production is notoriously challenging."*

- **Chirag Dekate**, Vice President Analyst, Gartner

From DevOps to MLOps

The ML process



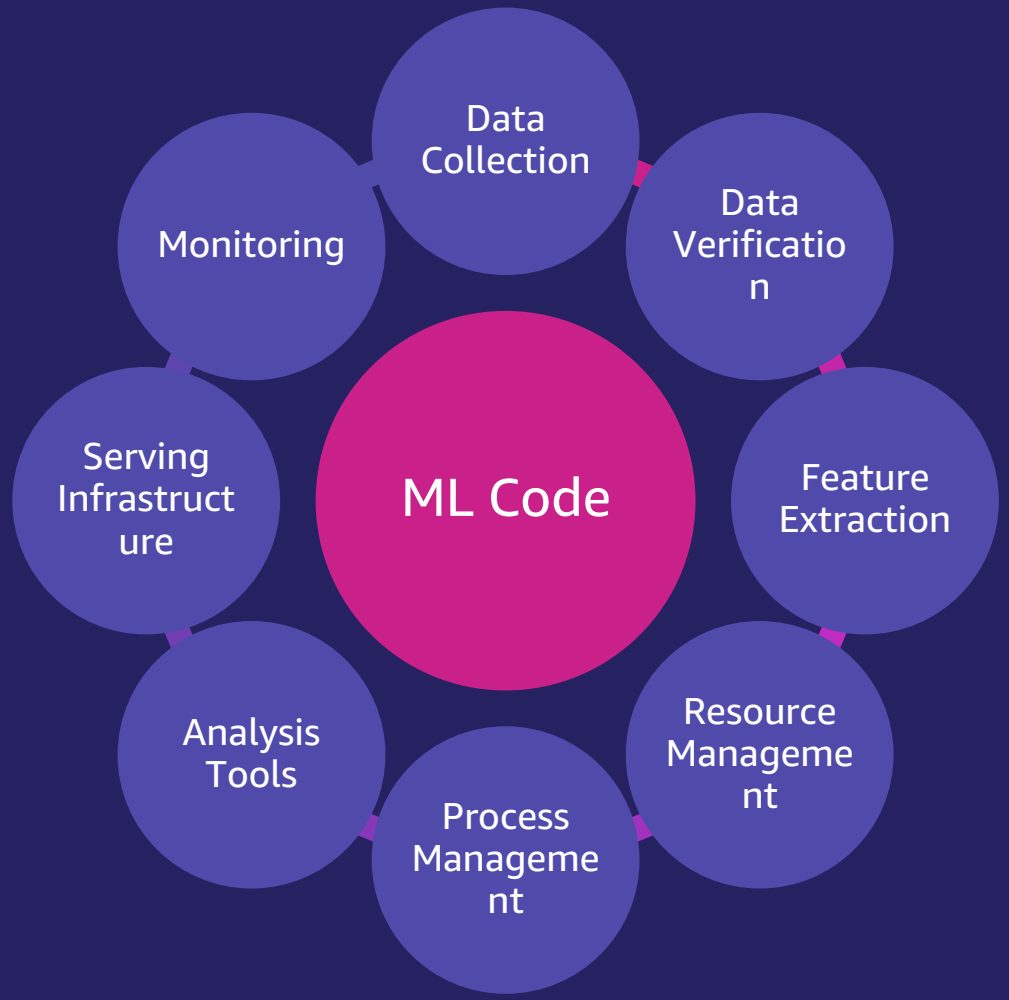
Phase 1: Research/Experiment

Question: "Can we use ML to solve this?"

- *"Is it possible to … ?"*
- *"Can we use this data to solve the following problem?"*
- *"Surely we must be able to …"*

Typical scenarios

- Scientific projects
- Proof-of-concepts (PoCs)



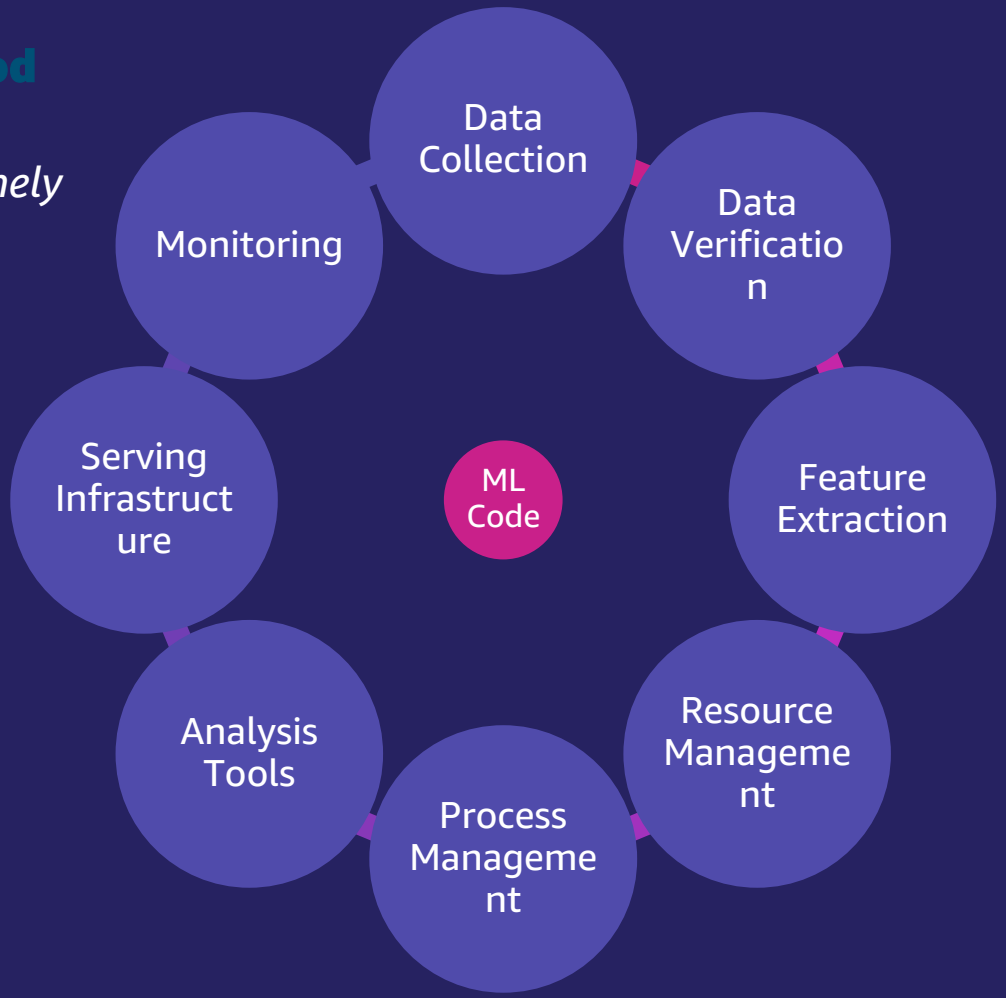
Phase 2: Operational

Question: "How do we implement this method at scale?"

- *How do we pipe the data into the model in a timely fashion?*
- *How do we collect, store and transform data so models can be retrained consistently?*
- *How do we build an A/B testing environment, in order to test future model iterations?*

Typical scenarios

- After PoC, bringing your ML models to production
- Migration of existing models into ML platform



MLOps – Why?

Agility

- Continuous and faster deliveries
- Faster modifications
- Faster bug-fixing

Experiments

- Faster and Controlled Experiments
- Faster integration of successful experiments to other environments

Scalability

- Ease integration of new ML model
- Standarization of code
- Lower operational costs

Time to Market

- Reduced time-to-market
- Faster planning and delivery expectations

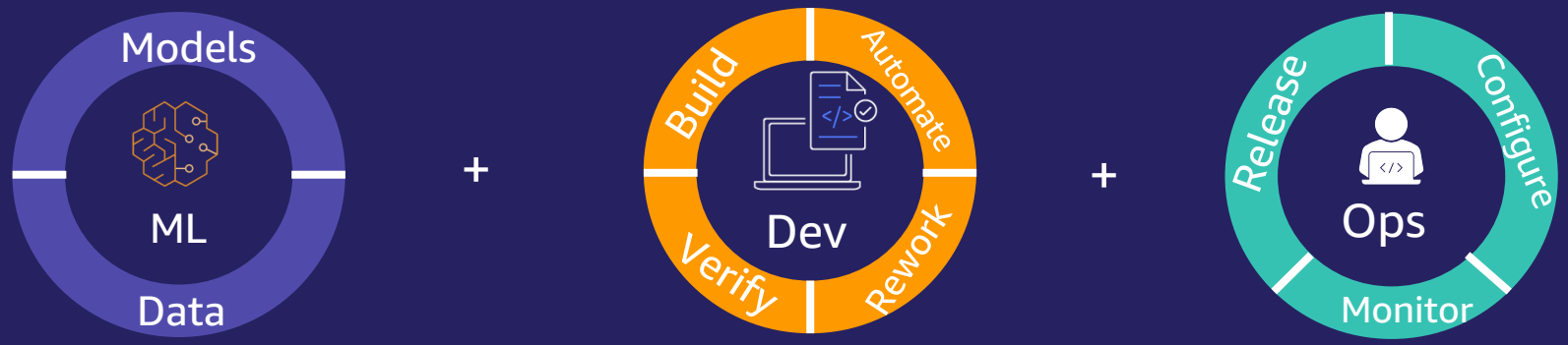
Business Owners

- Strong collaboration
- Improve iterations

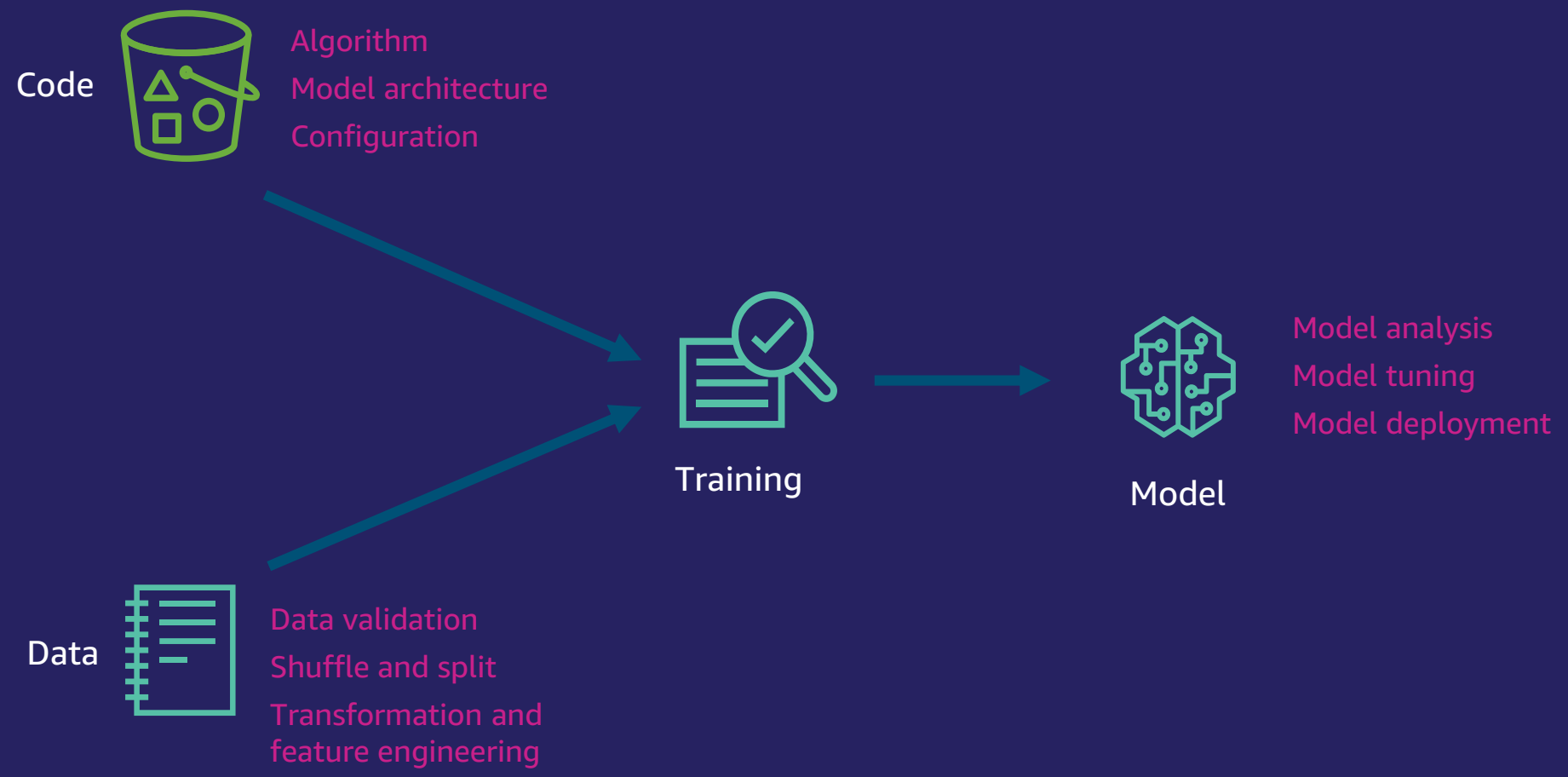
MLOps – What?

ML + Dev + Ops = MLOps

Collaborative and experimental in nature | Automate as much as possible | Continuous improvement of ML Models | Standardize and Scale



ML Code and Data are Independent



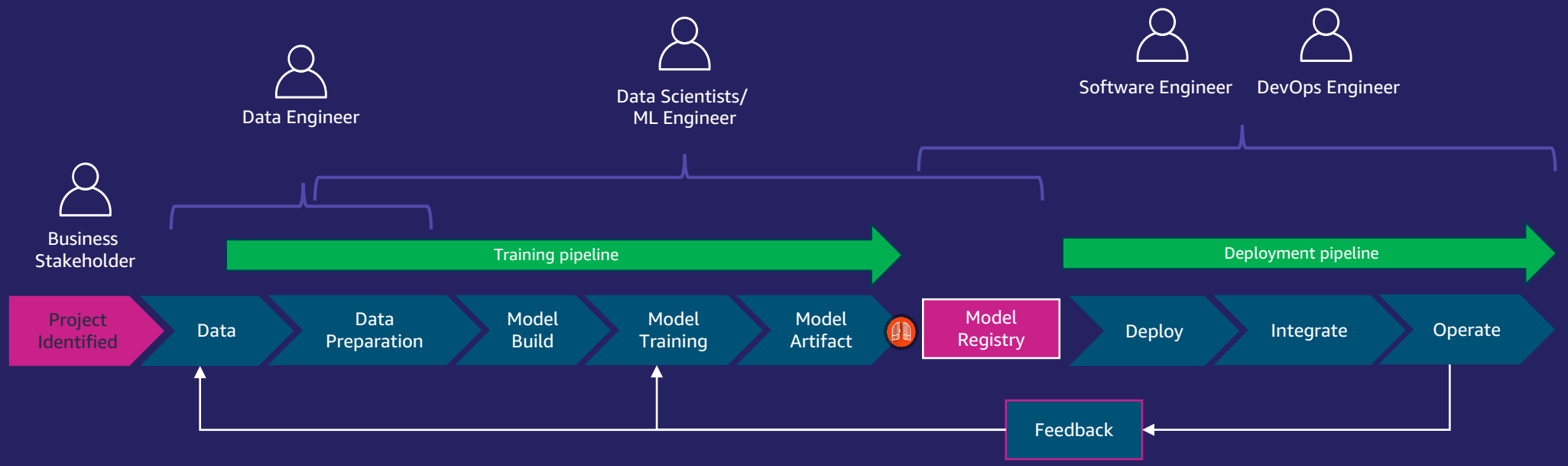
How is MLOps different from DevOps?

	DevOps	MLOPS
Code versioning	✓	✓
Compute environment	✓	✓
Continuous integration/delivery (CI/CD)	✓	✓
Monitoring in production	✓	✓
Data provenance		✓
Datasets		✓
Models		✓
Hyperparameters		✓
Metrics		✓
Workflows		✓

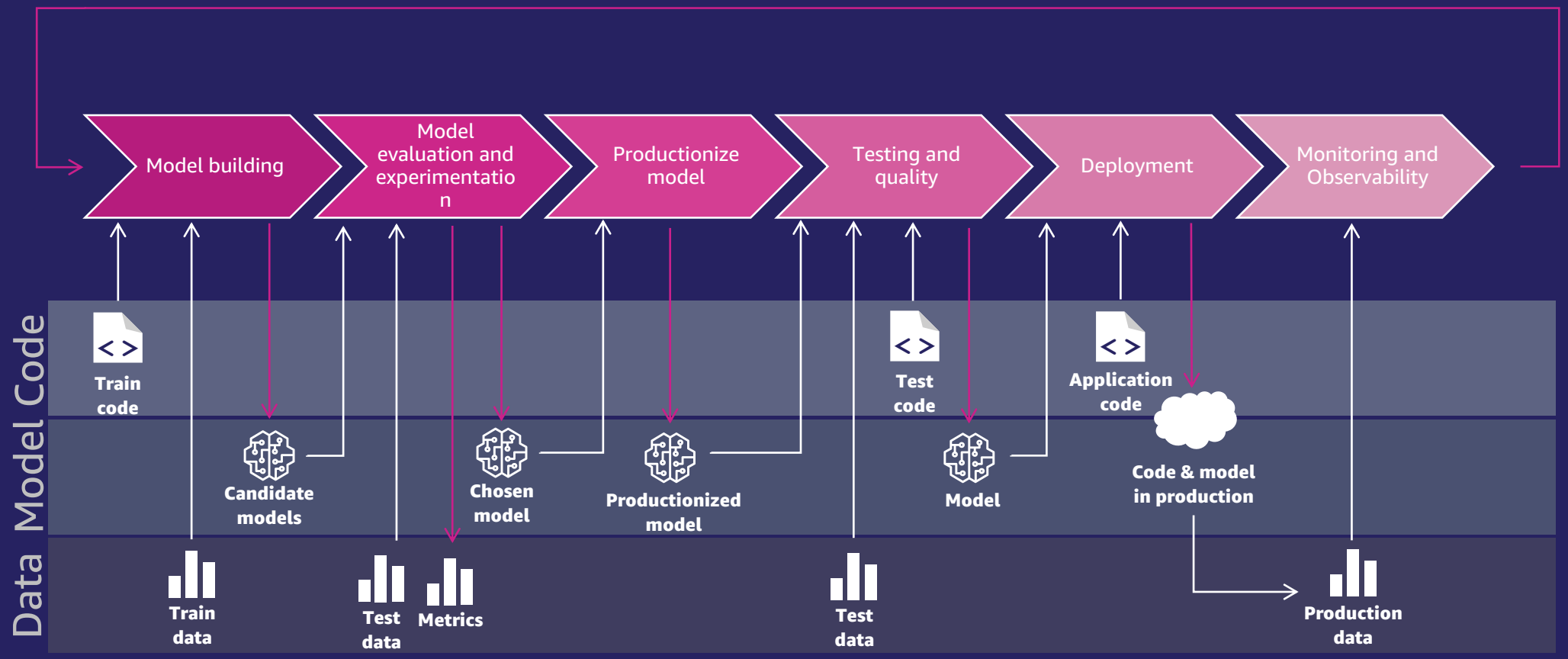
MLOPS
End-to-end ML lifecycle management

https://medium.com/analytics-vidhya/mlops-the-epoch-of-productionizing-ml-models-4eec06d93623

MLOps practices



ML lifecycle management

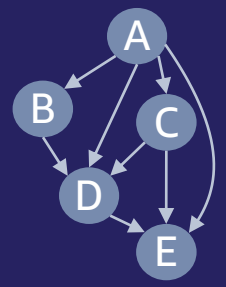


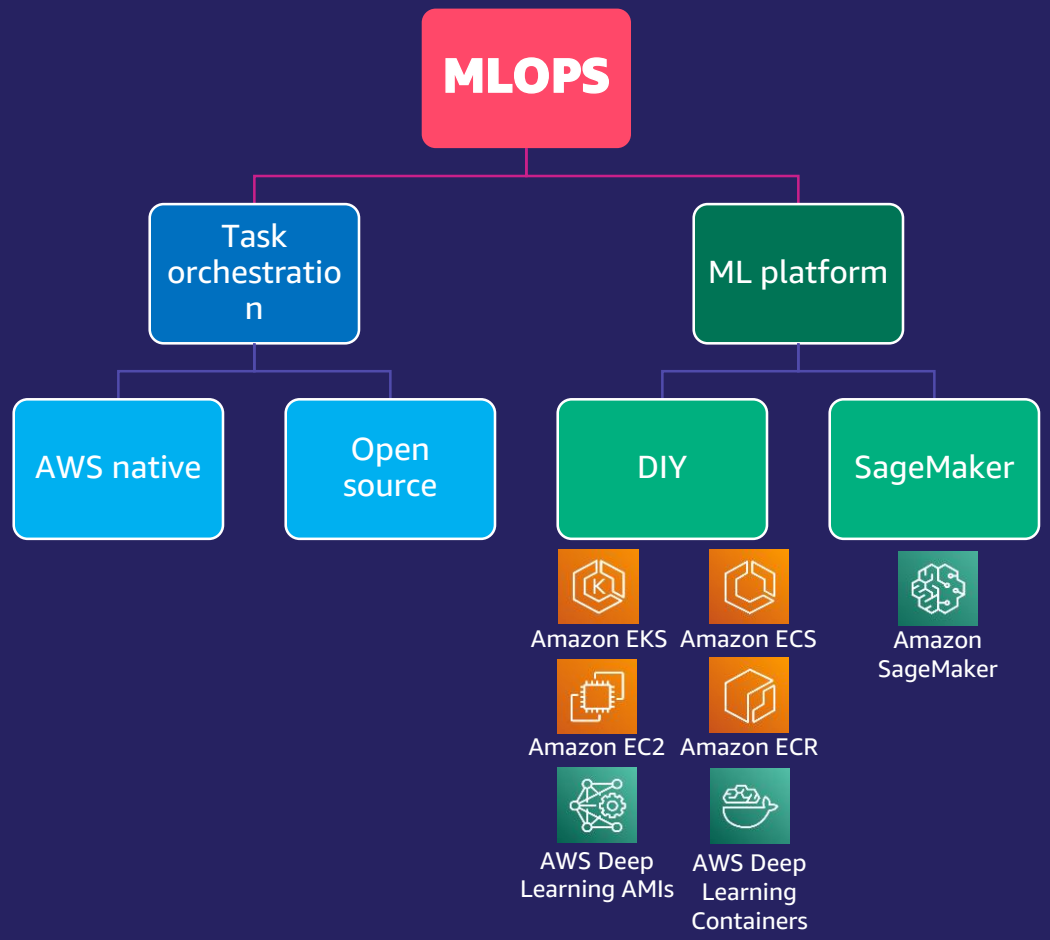
Automating ML Workflows

using SageMaker

Technology components in MLOps

- Create and manage workflows
- Automate ML steps & pipelines
- Implement CI/CD
- Form a Directed Acyclic Graph (DAG)





- ML development, experimentation, collaboration
- Compute/training environment
- Model registry
- Feature store
- Model deployment
- Monitoring in production
- Hyperparameter optimization
- Dataset management

Amazon SageMaker

Most complete, end-to-end ML service

Integrated Workbench

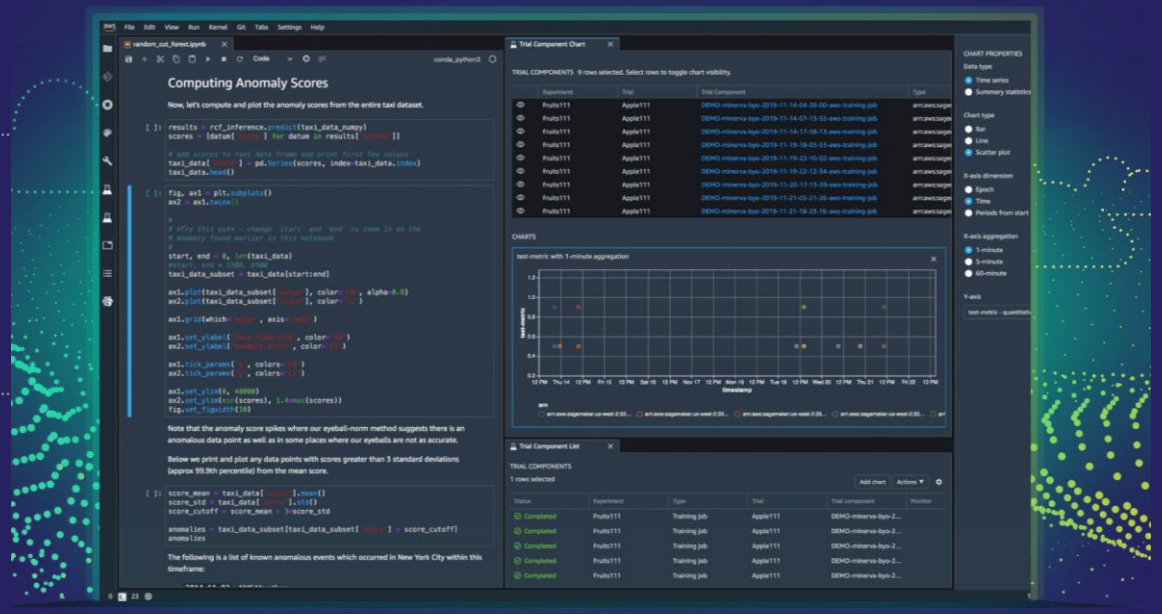
Capabilities designed specifically for ML, data preparation, experiment management, and workflows

Managed Infrastructure

Designed for ultra low latency and high throughput, automatic scaling, and distributed training

Managed Tooling

Purpose-built from the ground up to work together including auto ML, collaboration, debugger, profiler, bias analyzer, and explainability



Amazon SageMaker Features

Amazon SageMaker

PREPARE

SageMaker Ground Truth
Label training data for machine learning

SageMaker Data Wrangler
Aggregate and prepare data for machine learning

SageMaker Processing
Built-in Python, BYO R/Spark

SageMaker Feature Store
Store, update, retrieve, and share features

SageMaker Clarify
Detect bias and understand model predictions

BUILD

SageMaker Studio Notebooks
Jupyter notebooks with elastic compute and sharing

Built-in and Bring your-own Algorithms
Dozens of optimized algorithms or bring your own

Local Mode
Test and prototype on your local machine

SageMaker Autopilot
Automatically create machine learning models with full visibility

SageMaker JumpStart
Pre-built solutions for common use cases

TRAIN & TUNE

Managed Training
Distributed infrastructure management

SageMaker Experiments
Capture, organize, and compare every step

Automatic Model Tuning
Hyperparameter optimization

Distributed Training Libraries
Training for large datasets and models

SageMaker Debugger
Debug and profile training runs

Managed Spot Training
Reduce training cost by 90%

DEPLOY & MANAGE

Managed Deployment
Fully managed, ultra low latency, high throughput

Kubernetes & Kubeflow Integration
Simplify Kubernetes-based machine learning

Multi-Model Endpoints
Reduce cost by hosting multiple models per instance

SageMaker Model Monitor
Maintain accuracy of deployed models

SageMaker Edge Manager
Manage and monitor models on edge devices

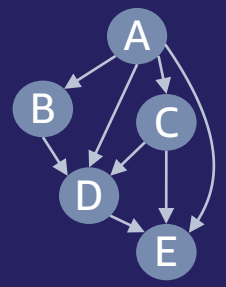
SageMaker Pipelines
Workflow orchestration and automation

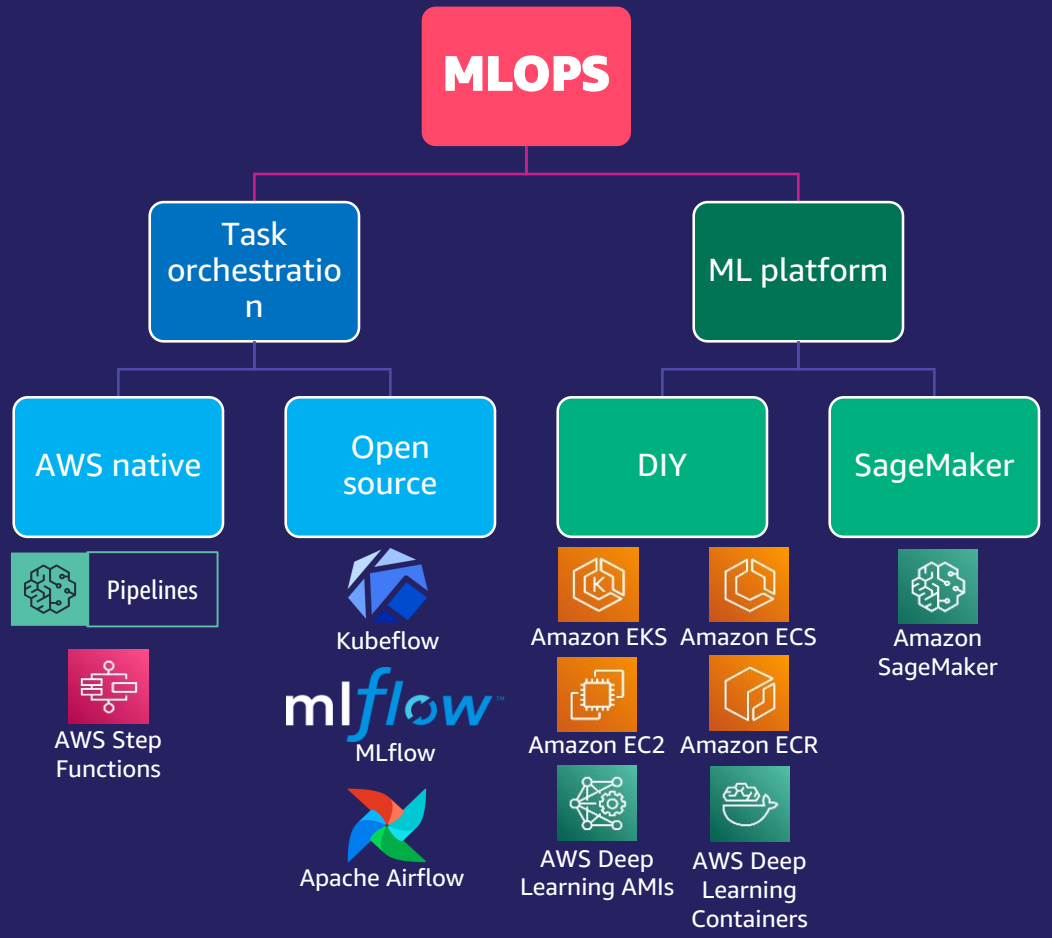
SageMaker Studio

Integrated development environment (IDE) for ML

Technology components in MLOps

- Create and manage workflows
- Automate ML steps & pipelines
- Implement CI/CD
- Form a Directed Acyclic Graph (DAG)





- ML development, experimentation, collaboration
- Compute/training environment
- Model registry
- Feature store
- Model deployment
- Monitoring in production
- Hyperparameter optimization
- Dataset management

Task orchestration

Open source 3rd party options

MLflow

Open source platform for the ML lifecycle

Apache Airflow

Platform to author, schedule and monitor workflows

Kubeflow

ML toolkit for Kubernetes

Native AWS options

AWS Step Functions

Serverless pipeline orchestration

Amazon SageMaker Pipelines

Managed ML pipelines in SageMaker Studio

Native integration with SageMaker

Apache Airflow

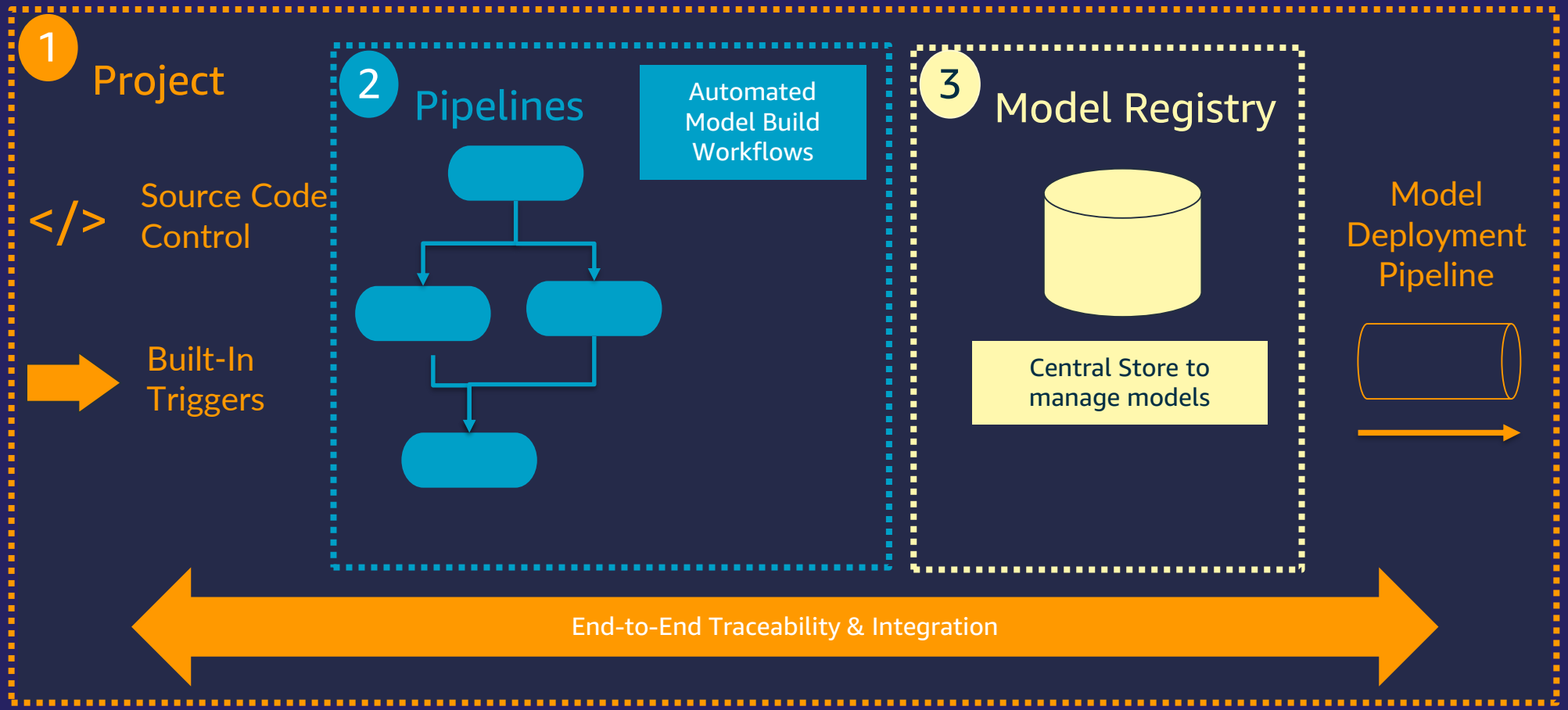
- SageMaker Operators in Apache Airflow
- Amazon Managed Workflows for Apache Airflow

(managed Airflow service)

Kubeflow & Kubernetes

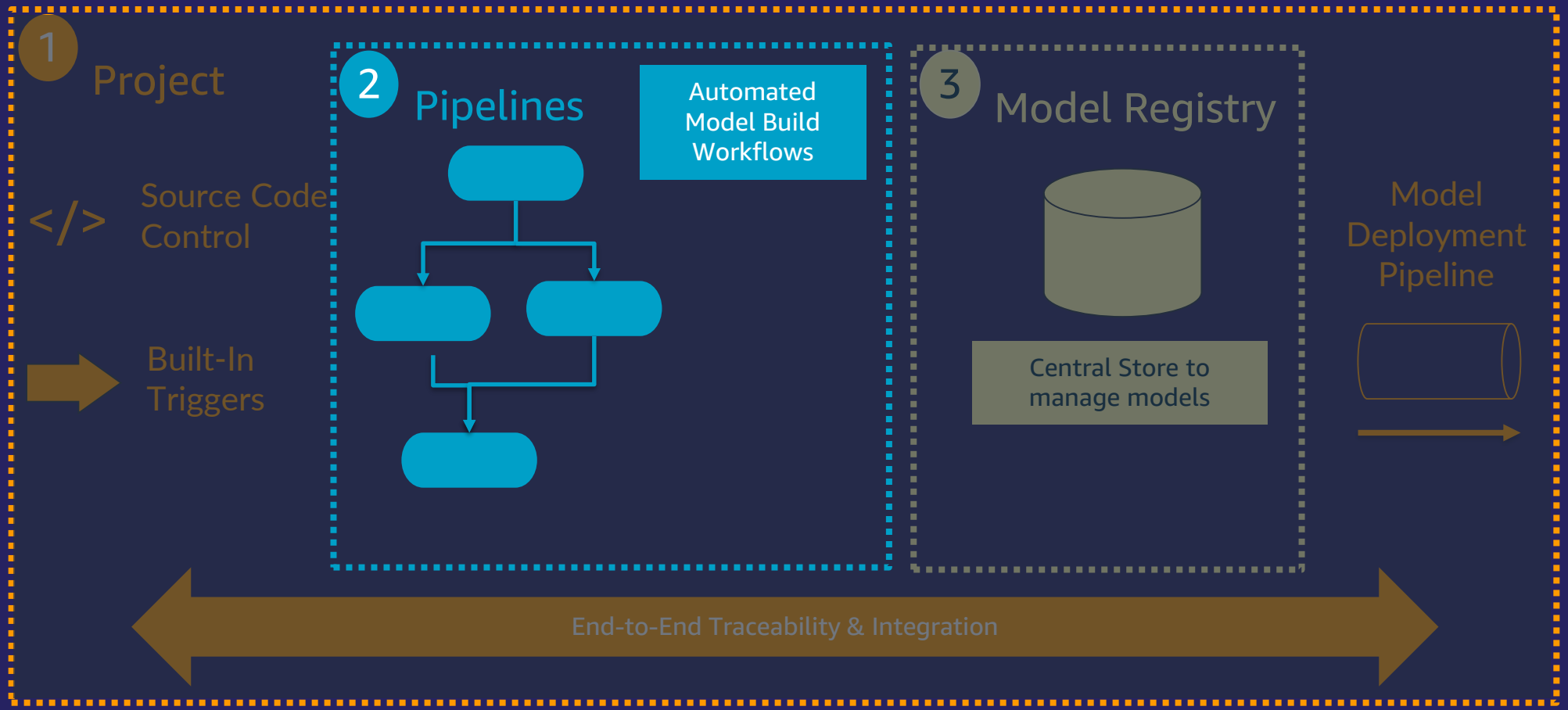
- SageMaker Components for Kubeflow Pipelines
- SageMaker Operators for Kubernetes

Amazon SageMaker Pipelines
Components



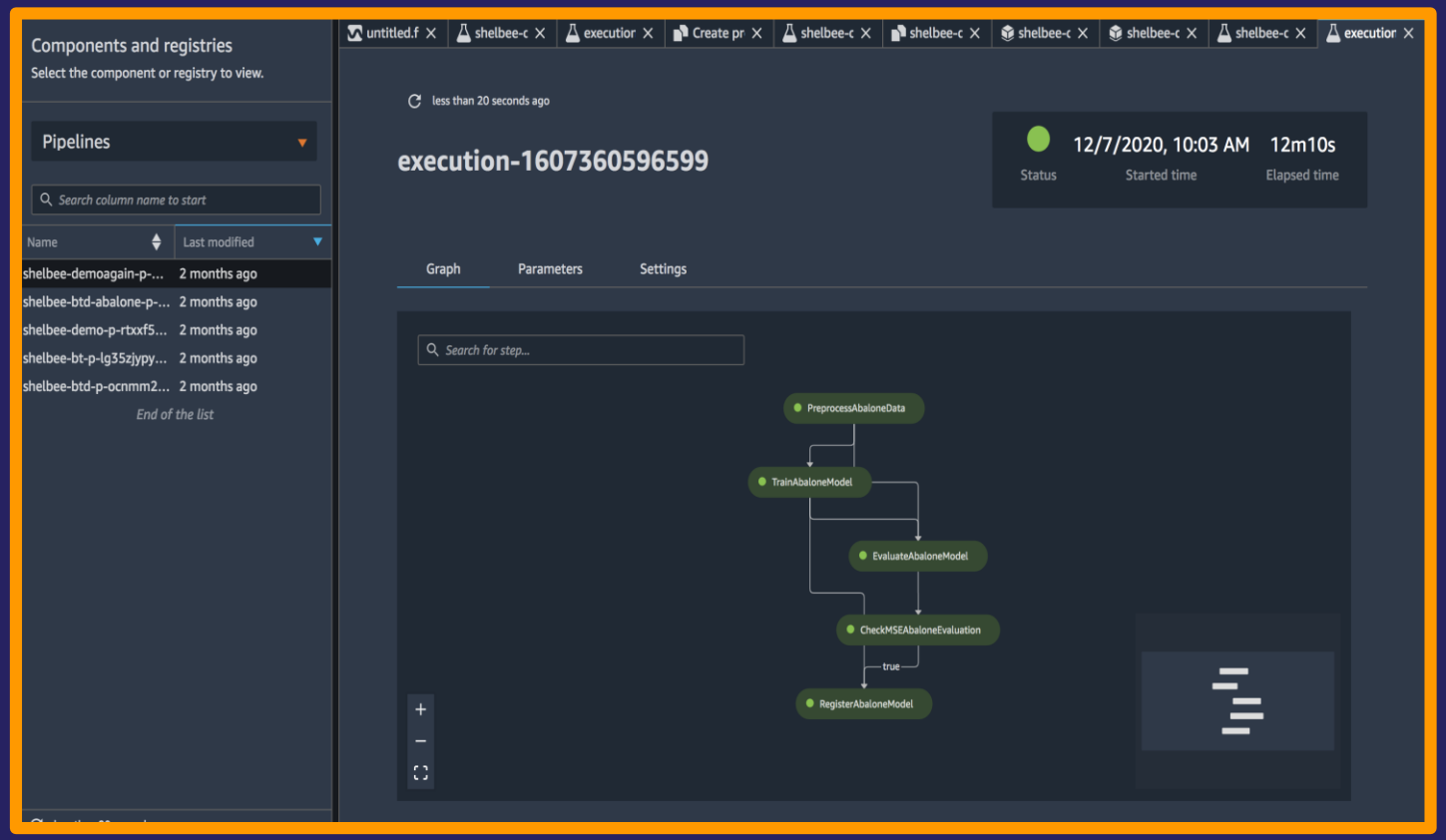
Amazon SageMaker Pipelines

Components – Pipelines



Amazon SageMaker Pipelines

Components – Pipelines

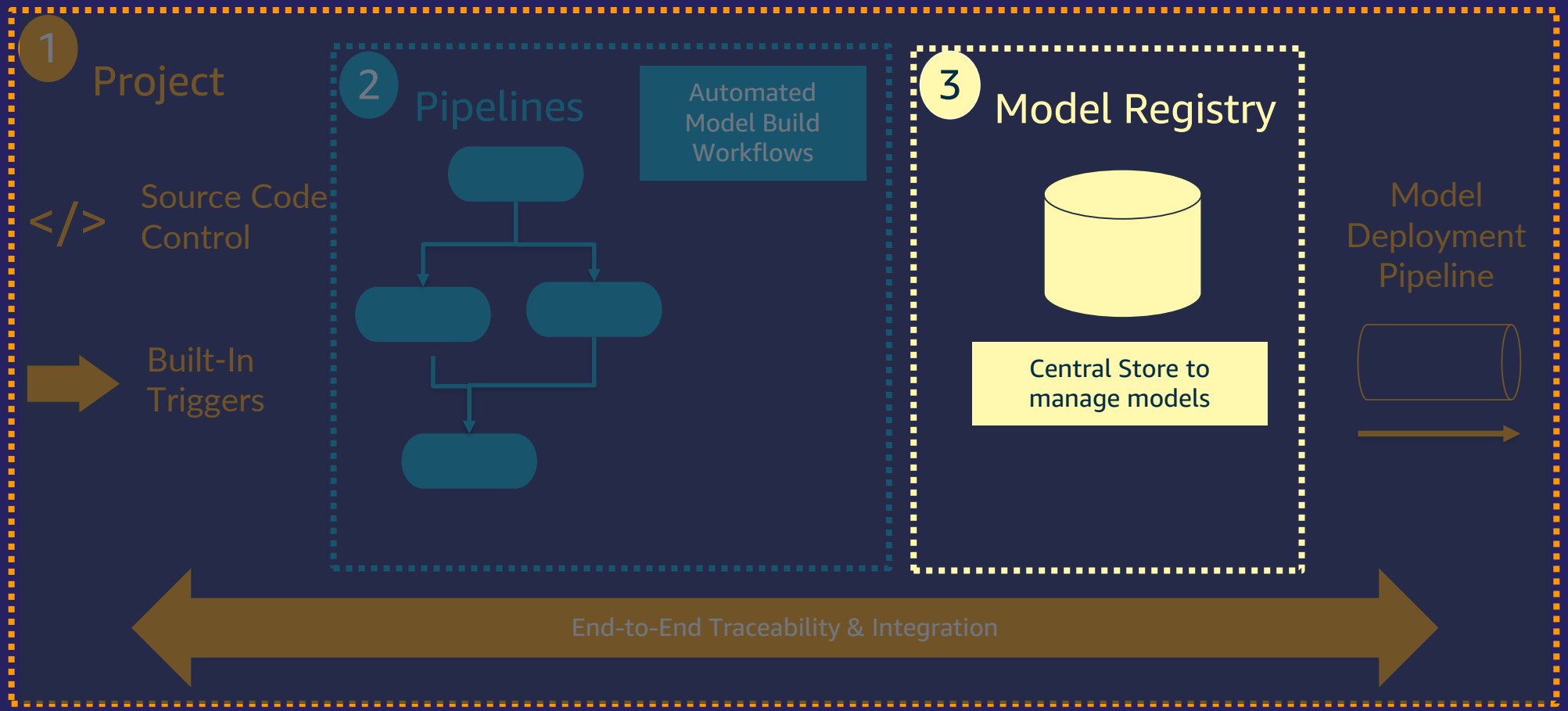


Supported Steps:

- Processing
- Training
- Tuning
- Conditional
- Register Model
- Create Model

Amazon SageMaker Pipelines

Components – Model Registry



Amazon SageMaker Pipelines

Components – Model Registry

Versions | Settings

Search column name to start

Version	Stage	Status	Short description	Modified by	Last modified
3	None	Pending			
2	prod	Approved		workshop-user	
1	staging	Approved		shelbee-iggy	

Version 2

Status	Pipeline	Execution	Last Stage	Model group	
Approved	shelbee-demoagain-p-...	workflow-2	prod	shelbee-demoagain-p-...	Update status

Activity | Metrics | Settings

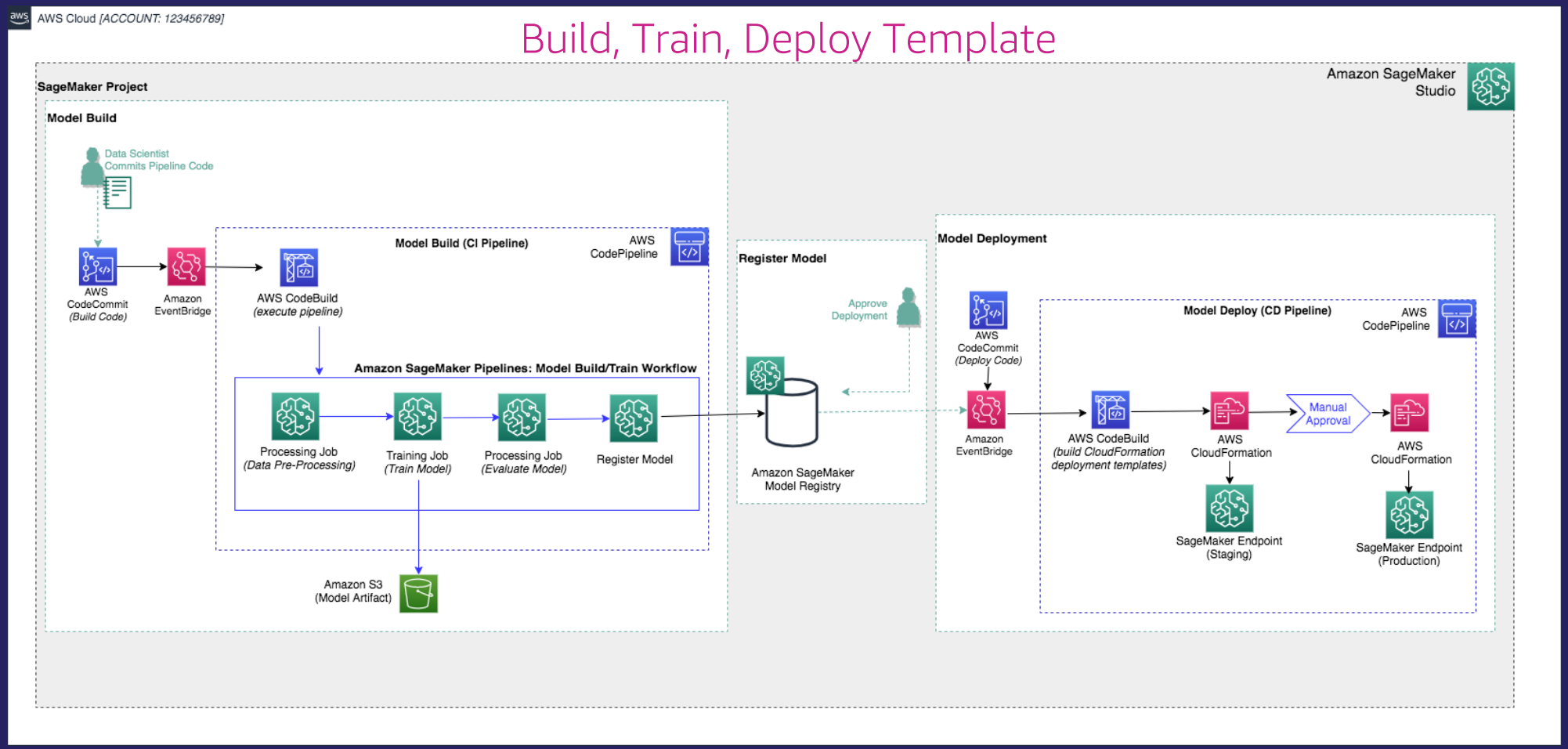
Model metrics

Model metric	Metric value	Standard deviation
mse	4.823176167079859	2.1960237865043672

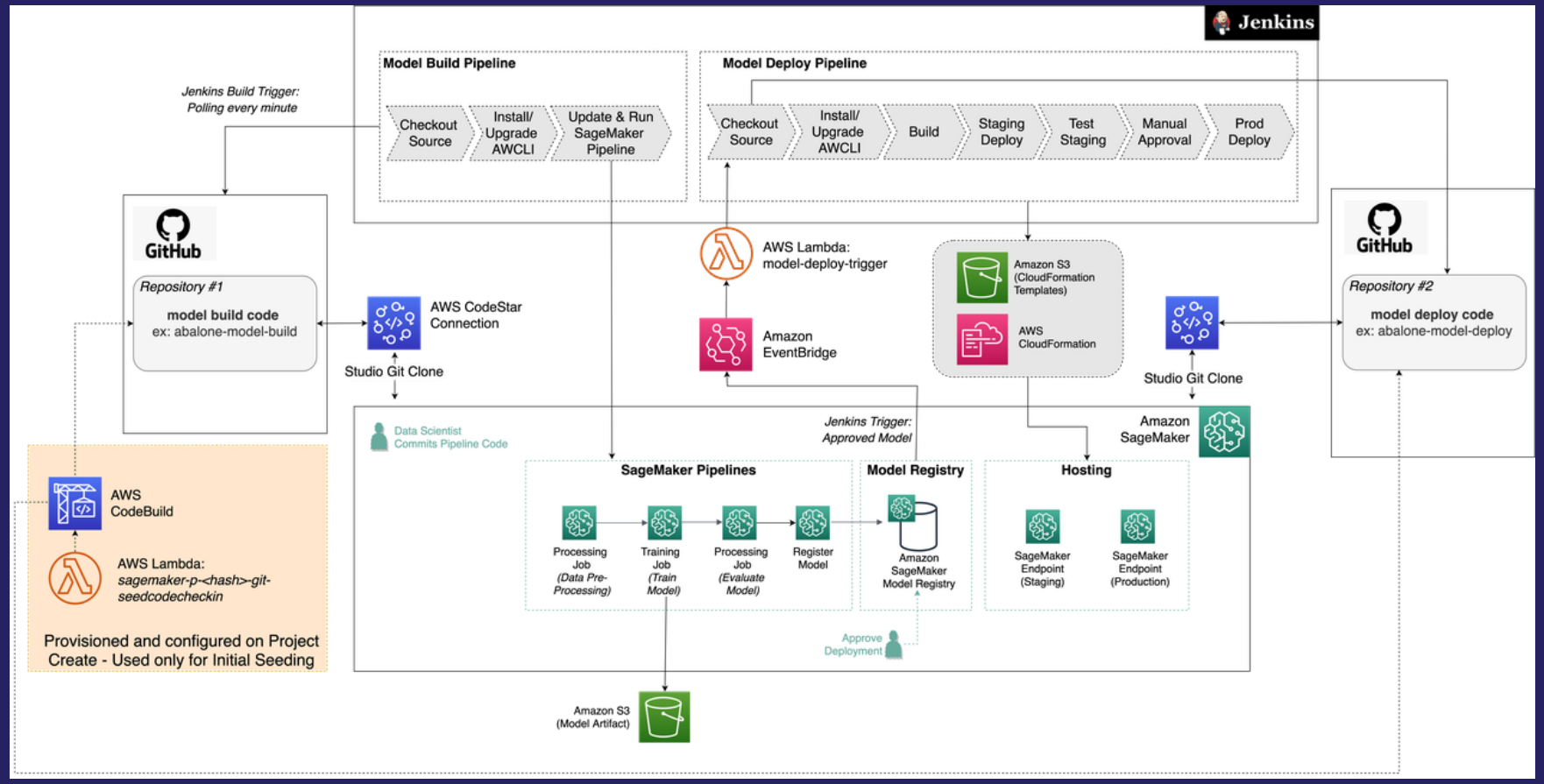
- Catalog models for production
- Manage model versions
- Associate metadata with a model
- Manage the approval status of a model
- Deploy models to production (with Projects)
- Track model performance metrics

Amazon SageMaker Projects

High Level Services View



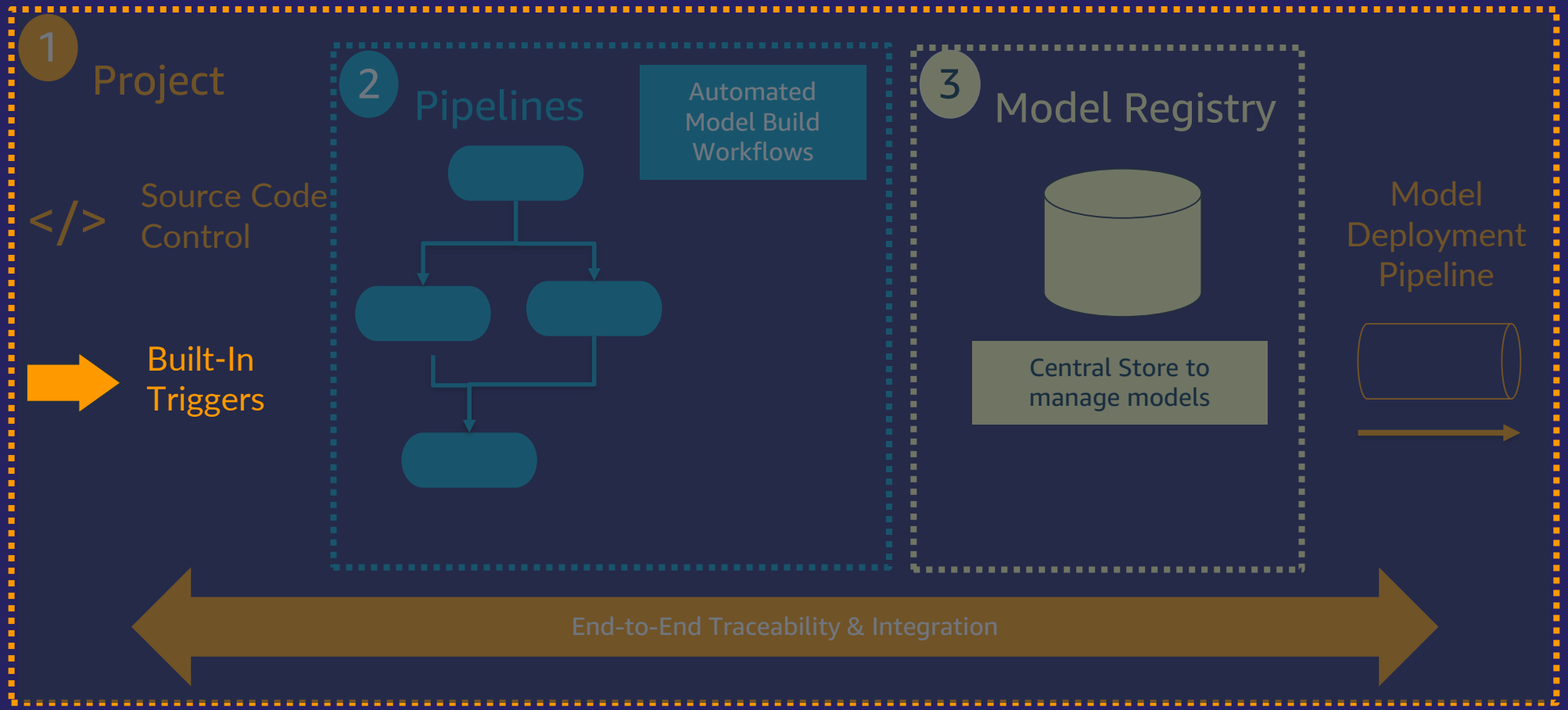
Amazon SageMaker Projects
using third-party source control and Jenkins



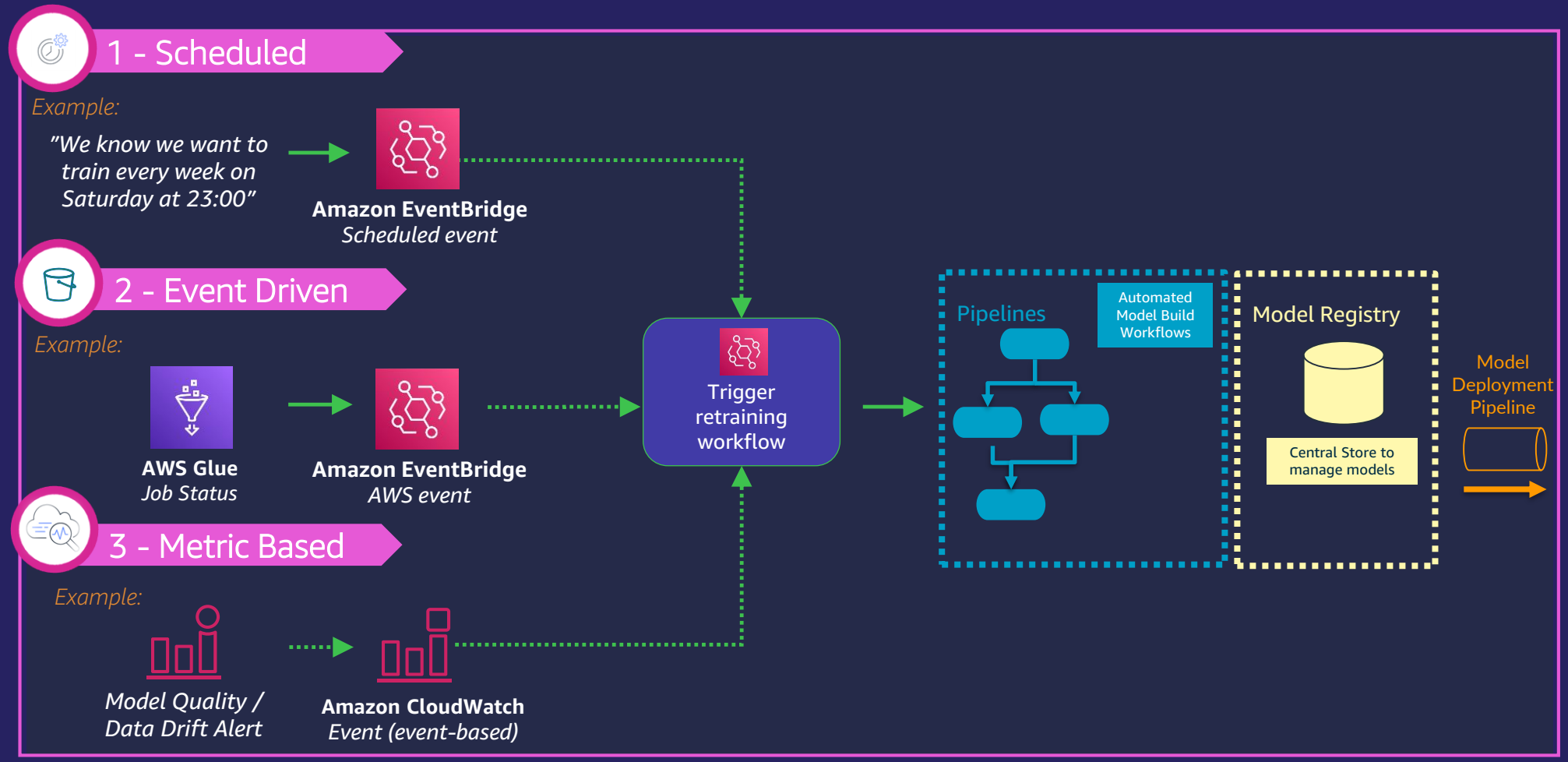
https://aws.amazon.com/blogs/machine-learning/create-amazon-sagemaker-projects-using-third-party-source-control-and-jenkins/

Amazon SageMaker Pipelines

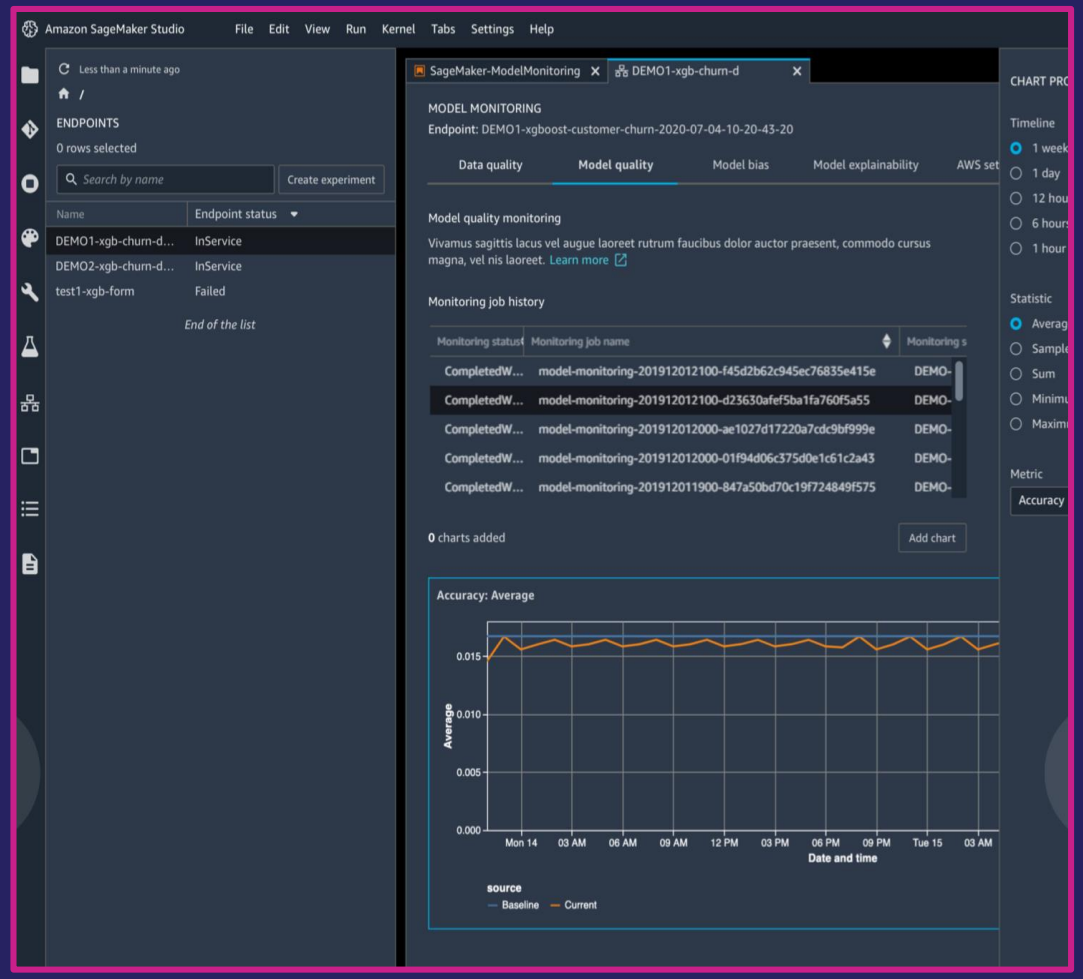
Built-In Triggers



Creating retraining strategies



Amazon SageMaker Model Monitor



Supported Features:

- Automatic data collection
- Continuous monitoring
- Flexible Monitoring Rules
- Visual data analysis
- CloudWatch integration

Amazon SageMaker MLOps

Streamline the ML lifecycle

- **Automate ML workflows to scale model development**
- **Build CI/CD pipelines for ML to accelerate model deployment**
- **Catalog model versions, metadata, metrics, and approvals for traceability and reusability**
- **Track lineage for troubleshooting and compliance**
- **Maintain accuracy of predictions after models are deployed**
- **Enhance governance and security**

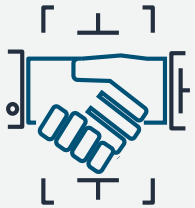
aws machine learning

Getting Started

Getting started: Next steps

Discovery and Get Hands Dirty

Proof of Concepts (PoC)



AWS Partner Network (APN)

Training and Certification

Thank you! Fill in the event survey and get USD 25 AWS Credits