AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

DOP211-R2

Building scalable machine learning pipelines

Shelbee Eigenbrode
Principal AI/ML Specialist Solutions Architect
Amazon Web Services

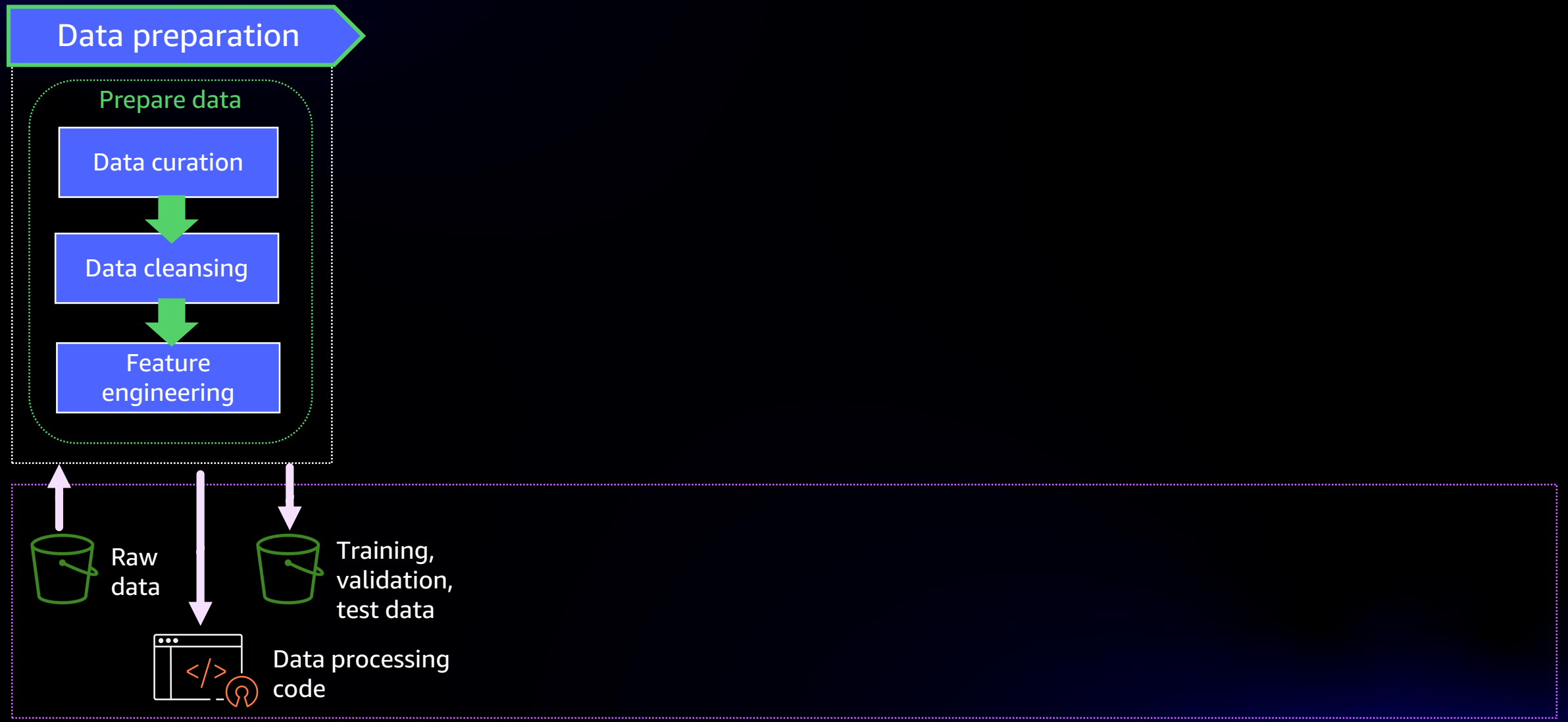
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Typical machine learning workflow
- Machine learning pipelines
- Amazon SageMaker Pipelines deep dive

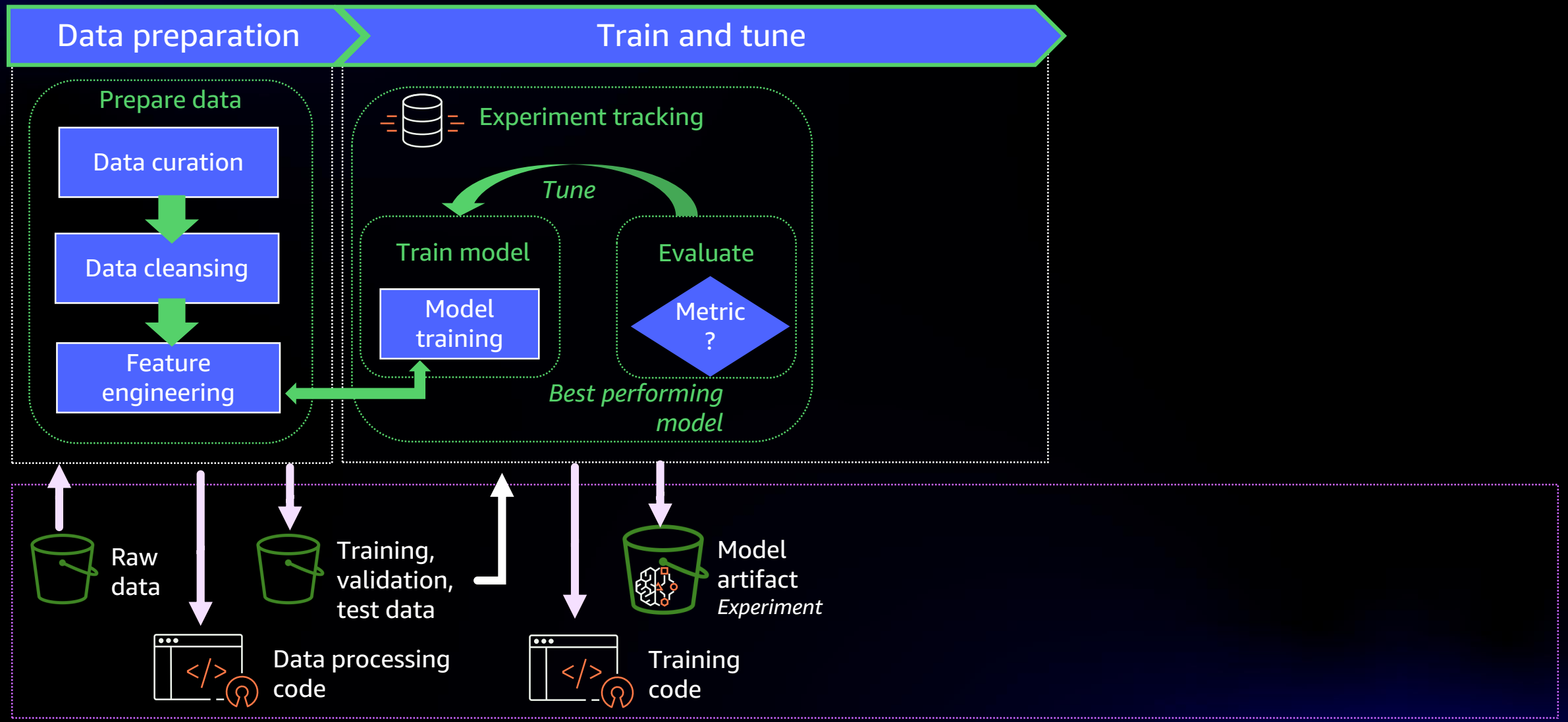
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning workflow



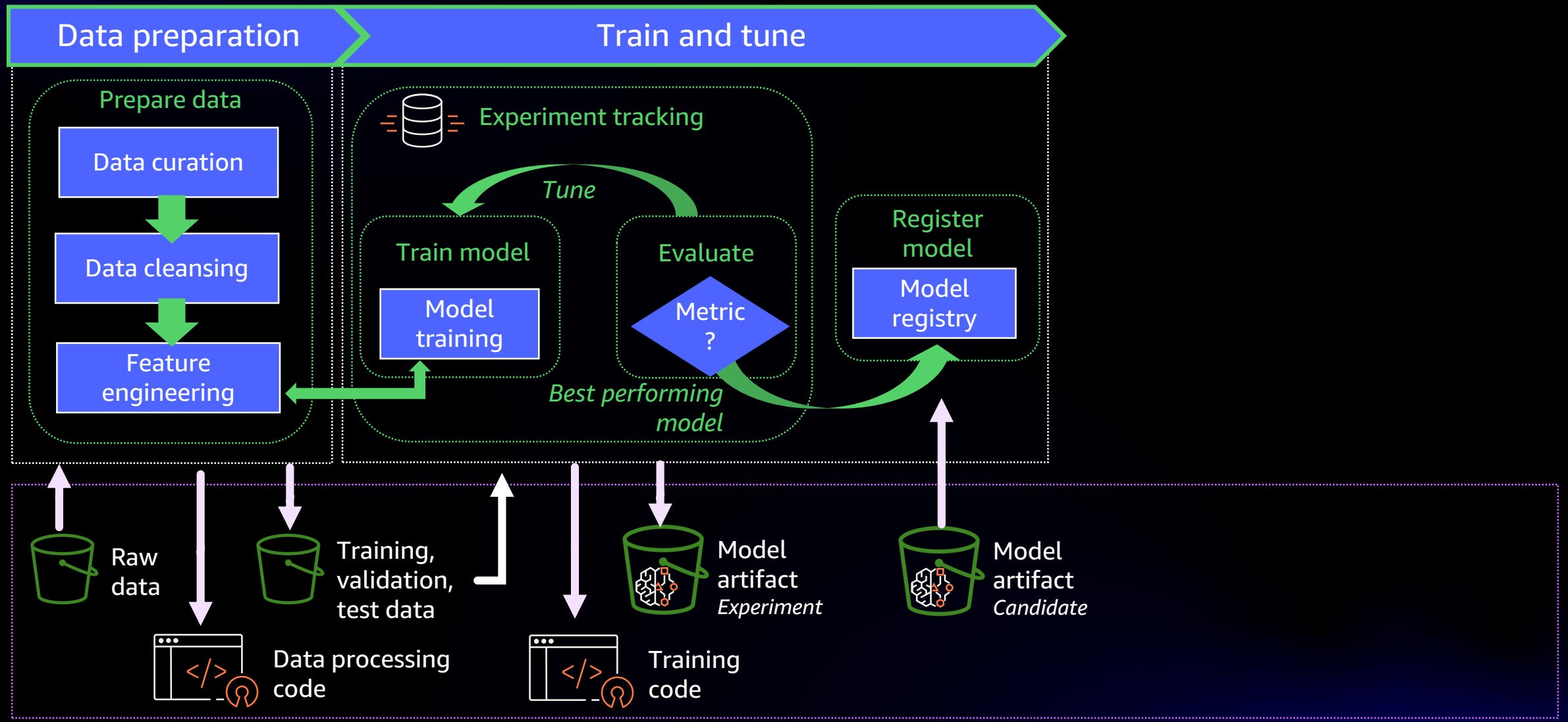
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning workflow



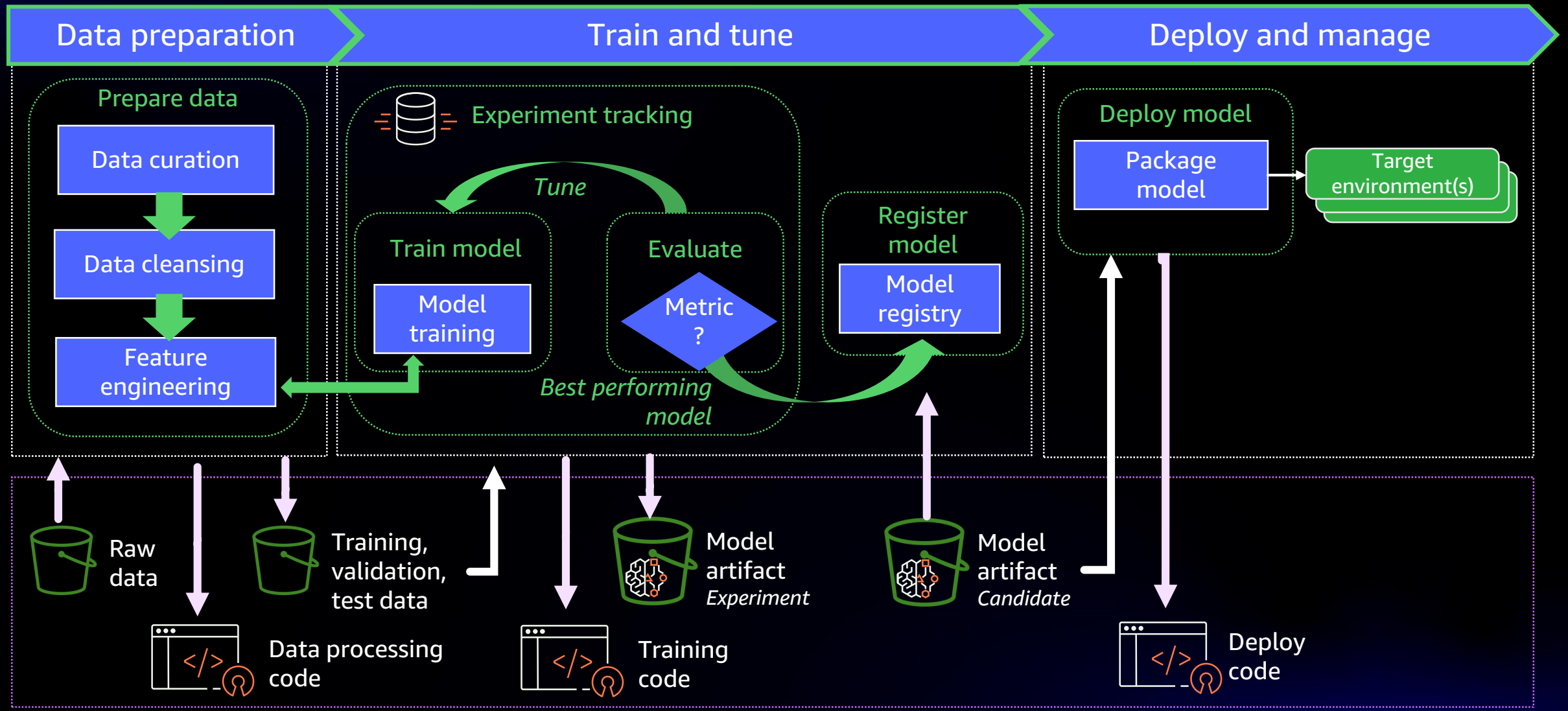
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning workflow



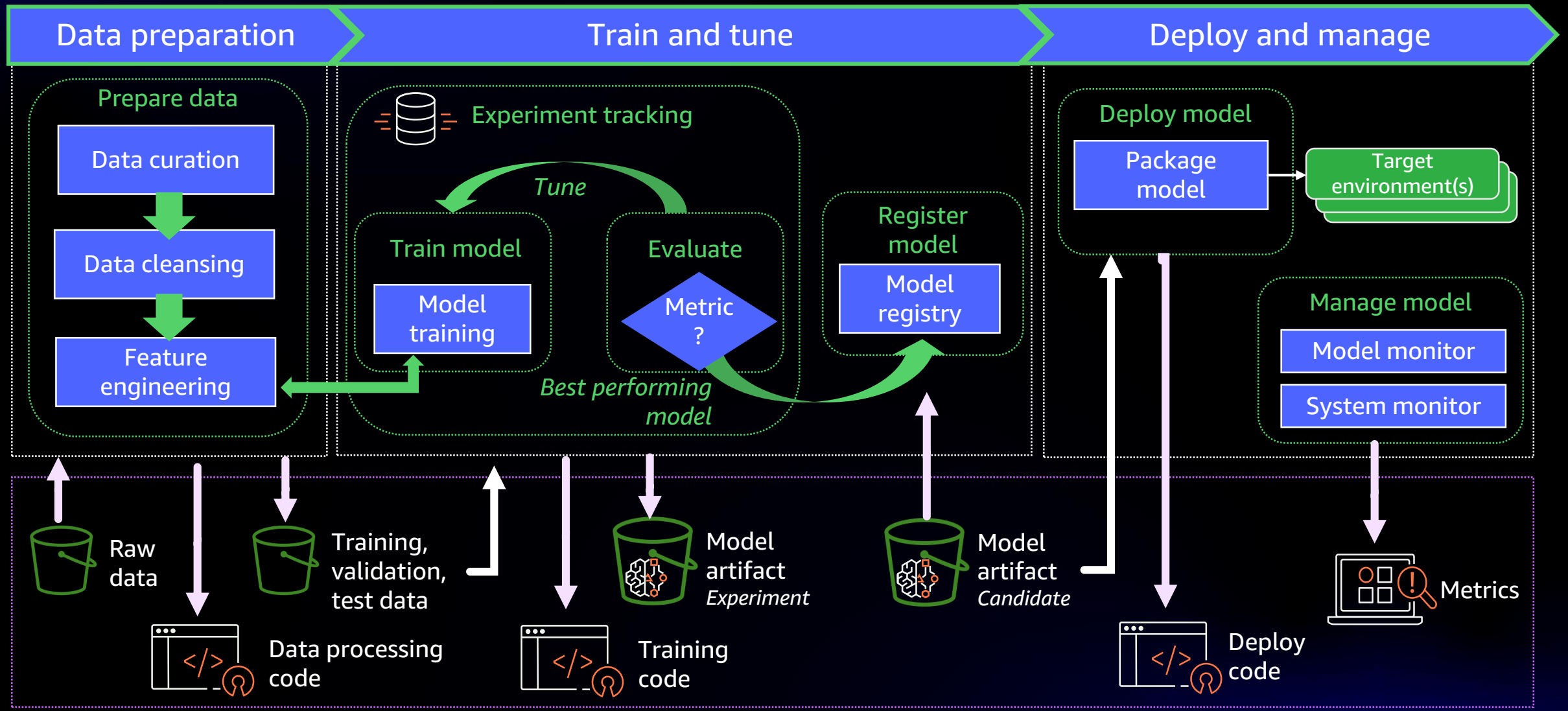
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning workflow



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning workflow



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning workflow



Challenges in building machine learning pipelines ➔

- Machine learning development lifecycle != software development lifecycle
- Multiple disparate pipelines
  - ✓ Data curation
  - ✓ Data preparation
  - ✓ Model training and tuning
  - ✓ Model deployment

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

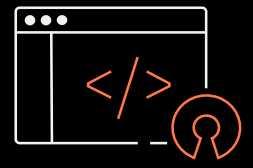
Machine learning pipelines

SERVICES AND TOOLING →

Purpose-built services for automating machine learning tasks



Purpose-built services for incorporating everything-as-code and CI/CD practices



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ML pipelines: The Golden Pipeline



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning pipelines

SERVICES AND TOOLING →

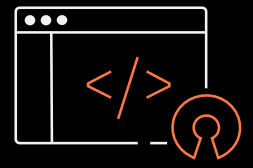
Purpose-built services for automating machine learning tasks



AWS-native examples:

- Amazon SageMaker Pipelines
- Amazon Managed Workflows for Apache Airflow (Amazon MWAA)
- AWS Step Functions

Purpose-built services for incorporating everything-as-code and CI/CD practices



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning pipelines

SERVICES AND TOOLING →

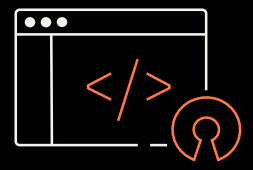
Purpose-built services for automating machine learning tasks



AWS-native examples:

- Amazon SageMaker Pipelines
- Amazon Managed Workflows for Apache Airflow (Amazon MWAA)
- AWS Step Functions

Purpose-built services for incorporating everything-as-code and CI/CD practices



AWS-native examples:

- Amazon SageMaker Projects
- AWS CodeBuild
- AWS CodeCommit
- AWS CodePipeline
- AWS CloudFormation

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning pipelines

SERVICES AND TOOLING →

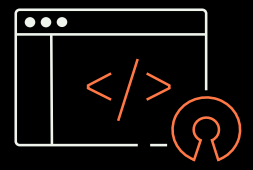
Purpose-built services for automating machine learning tasks



AWS-native examples:

- Amazon SageMaker Pipelines
- Amazon Managed Workflows for Apache Airflow (Amazon MWAA)
- AWS Step Functions

Purpose built services for incorporating everything-as-code and CI/CD practices



AWS-native examples:

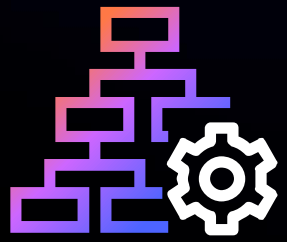
- Amazon SageMaker Projects
- AWS CodeBuild
- AWS CodeCommit
- AWS CodePipeline
- AWS CloudFormation

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker Pipelines deep dive

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Pipelines



Python SDK for quickly and easily building automated ML workflows



Catalog models to manage models at scale and trigger automated deployment workflows



Built-in support for CI/CD and end-to-end lineage tracking

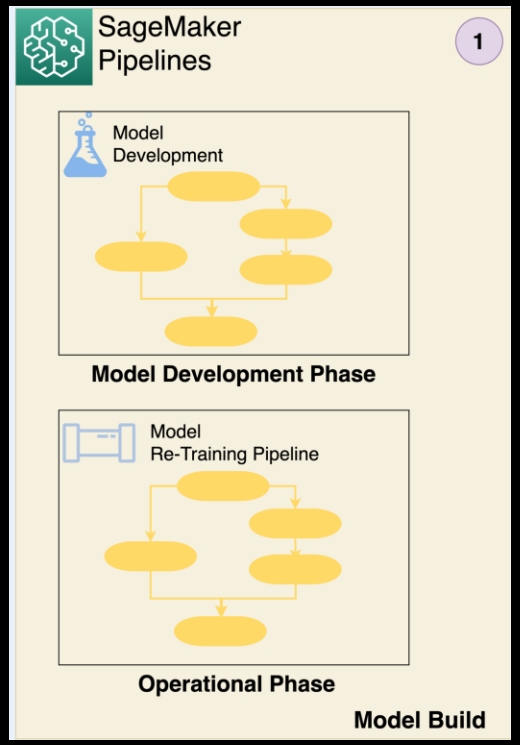
aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Pipelines

3 core components

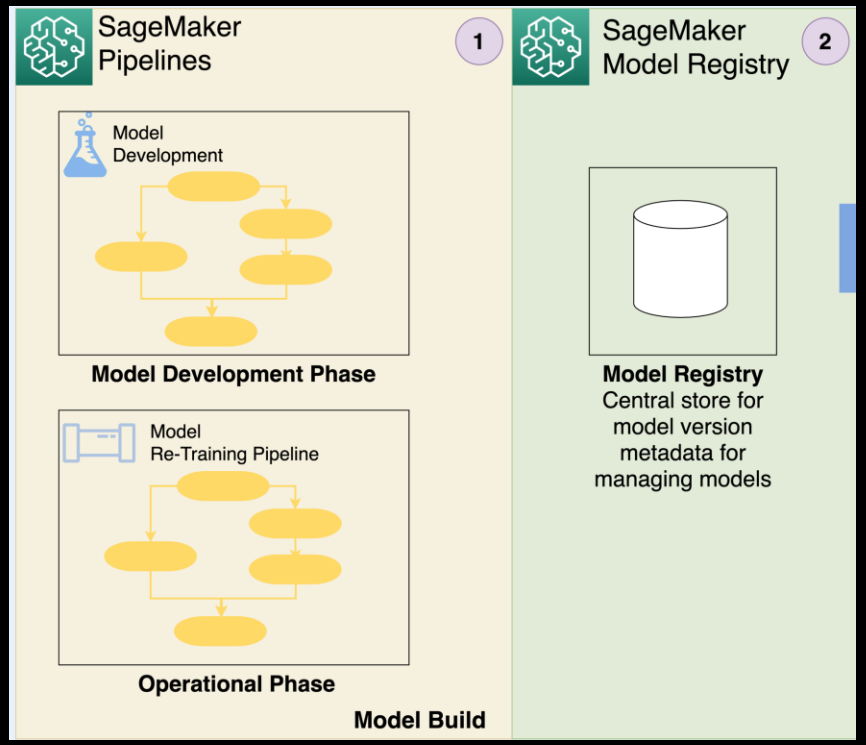
1. Pipelines



Amazon SageMaker Pipelines

3 core components

1. Pipelines
2. Model Registry

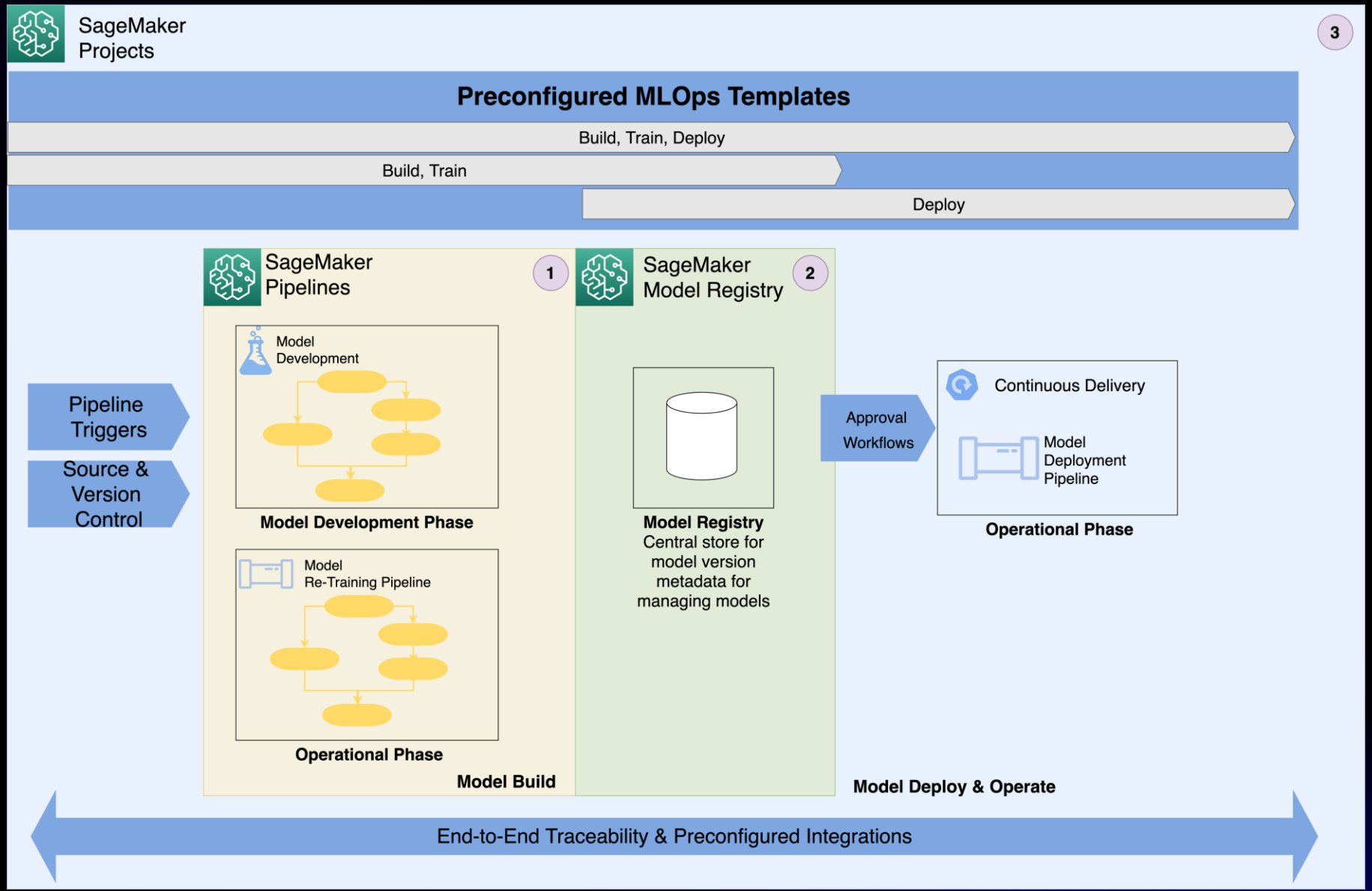


aws

Amazon SageMaker Pipelines

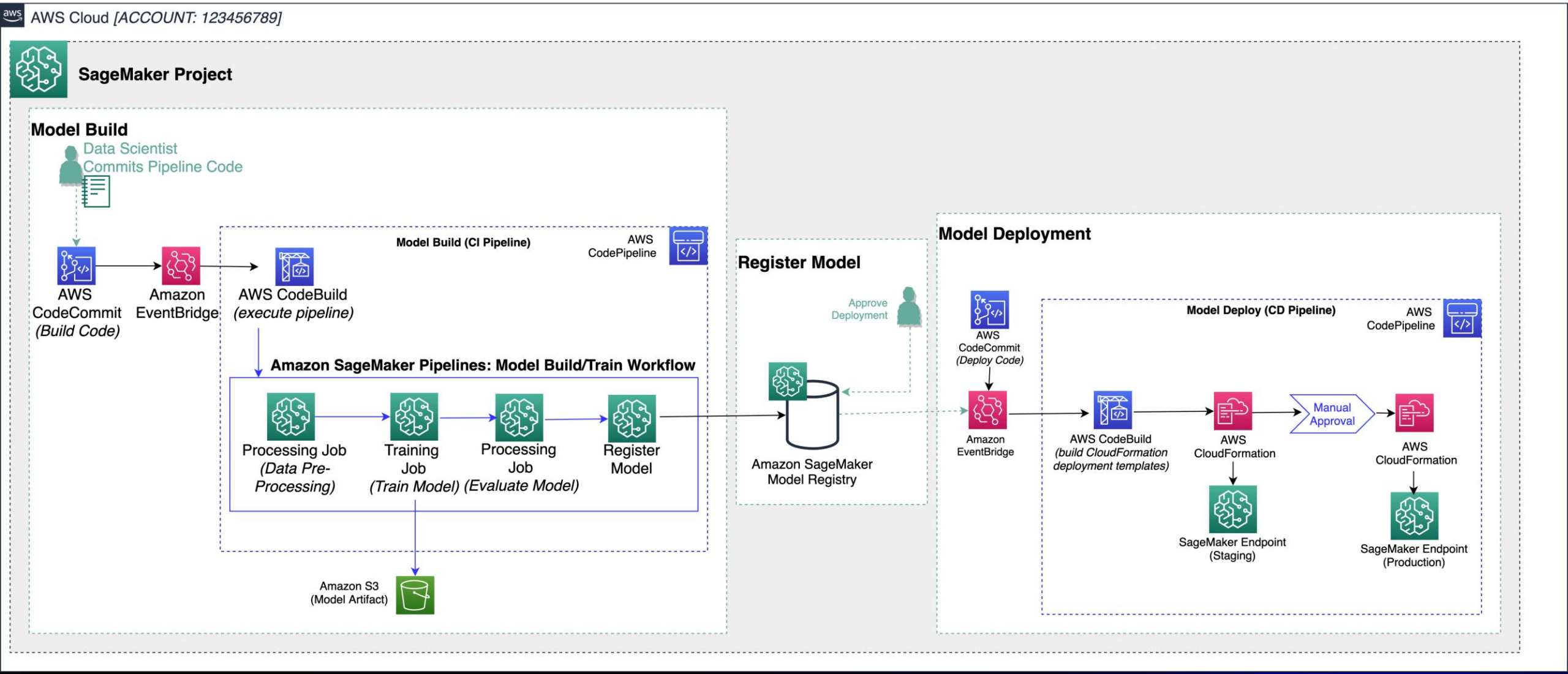
3 core components

1. Pipelines
2. Model Registry
3. Projects



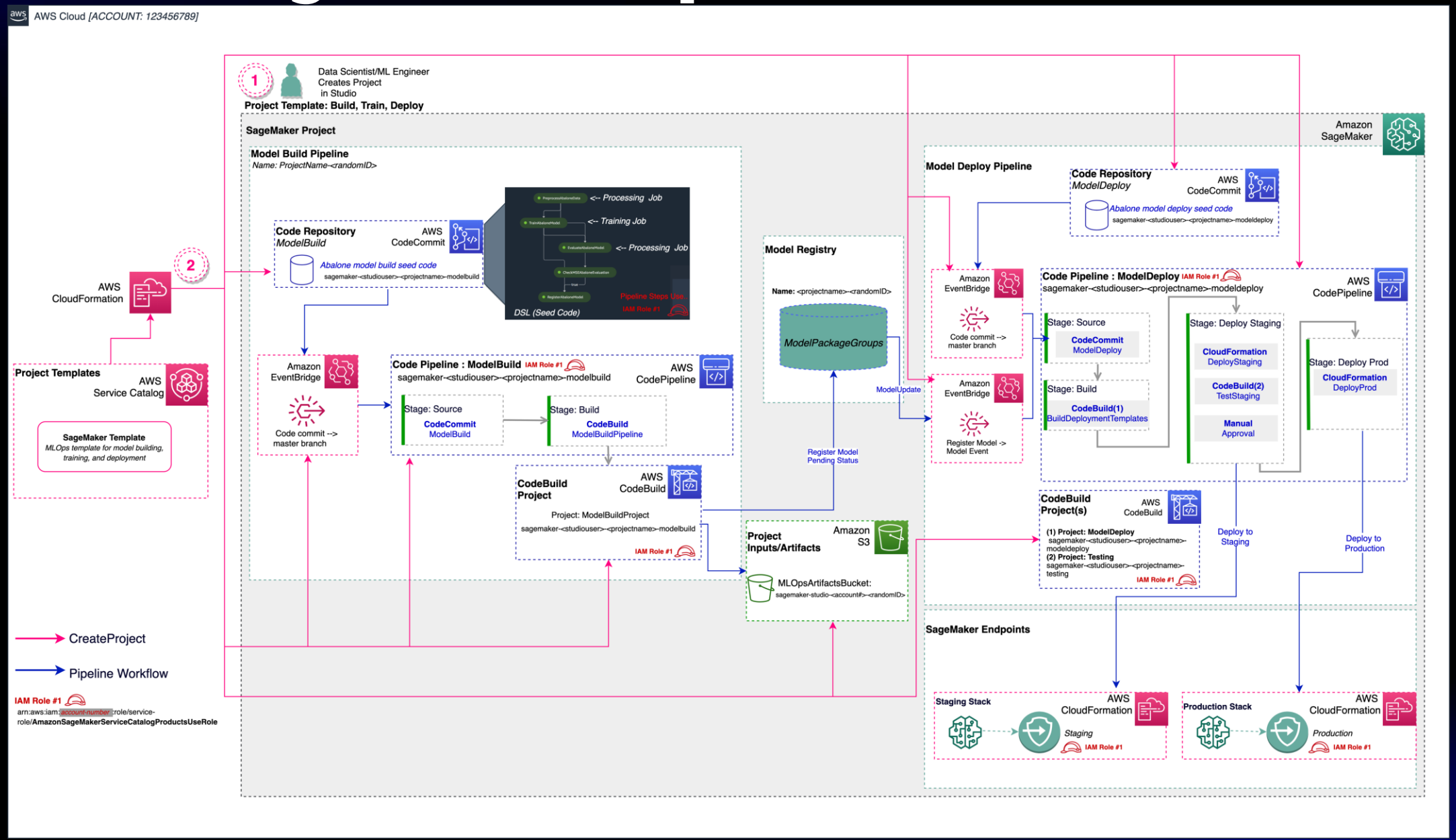
Amazon SageMaker Pipelines

BUILT-IN MLOPS PROJECT TEMPLATE: BUILD, TRAIN, DEPLOY



DEEP DIVE VIEW

Amazon SageMaker Pipelines



Amazon SageMaker Pipelines

BUILDING CUSTOM MLOPS PROJECT TEMPLATES

1 *Create your custom template*



CloudFormation

Required parameters to include:

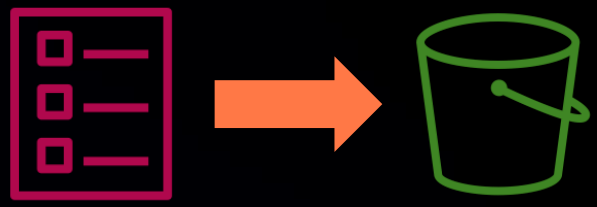
```
SageMakerProjectName:
    Type: String
    Description: Name of the project

  SageMakerProjectId:
    Type: String
    Description: Service generated Id of the project.
```



Tip: Use the built-in SageMaker templates as a starting point

2 *Upload CloudFormation template to Amazon S3*



Amazon S3

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

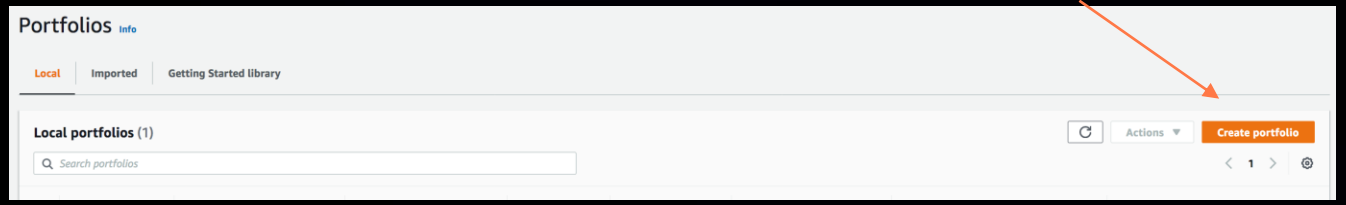
Amazon SageMaker Pipelines

BUILDING CUSTOM MLOPS PROJECT TEMPLATES

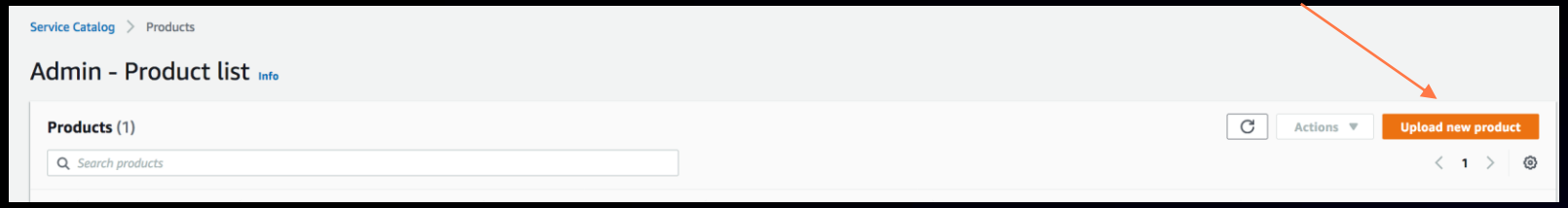
Service Catalog tasks

AWS Service Catalog

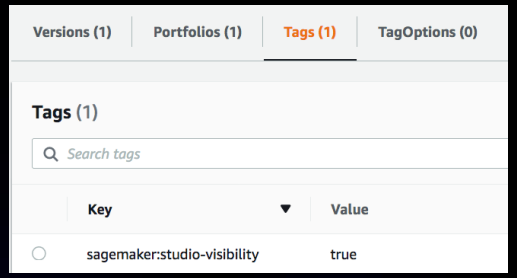
3 *Create portfolio*



4 *Create product*



Required tags



Your template will now be available for use in SageMaker Studio!

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning pipelines

GENERAL TIPS AND RECOMMENDATIONS →

Identify the inputs and outputs/artifacts of each step in your pipeline for version control

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning pipelines

GENERAL TIPS AND RECOMMENDATIONS →

Identify the inputs and outputs/artifacts of each step in your pipeline for version control

Implement a Model Registry for management of model versions

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning pipelines

GENERAL TIPS AND RECOMMENDATIONS →

Identify the inputs and outputs/artifacts of each step in your pipeline for version control

Implement a Model Registry for management of model versions

Incorporate automated quality gates to reduce manual handoffs between personas

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning pipelines

GENERAL TIPS AND RECOMMENDATIONS →

Identify the inputs and outputs/artifacts of each step in your pipeline for version control

Implement a Model Registry for management of model versions

Incorporate automated quality gates to reduce manual handoffs between personas

Use managed services to increase operational efficiencies

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning pipelines

GENERAL TIPS AND RECOMMENDATIONS →

Identify the inputs and outputs/artifacts of each step in your pipeline for version control

Implement a Model Registry for management of model versions

Incorporate automated quality gates to reduce manual handoffs between personas

Use managed services to increase operational efficiencies

Identify your model retraining strategy

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning pipelines

GENERAL TIPS AND RECOMMENDATIONS →

Identify the inputs and outputs/artifacts of each step in your pipeline for version control

Implement a Model Registry for management of model versions

Incorporate automated quality gates to reduce manual handoffs between personas

Use managed services to increase operational efficiencies

Identify your model retraining strategy

Incorporate steps supporting model monitoring into your pipeline (e.g., data baseline)

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Shelbee Eigenbrode

https://www.linkedin.com/in/shelbee-eigenbrode-5b632414/

Twitter: @Shelbee_SE
Email: shelbees@amazon.com

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.