aws machine learning

Architecting for ML, On AWS

Day 3

Phi Nguyen
Machine Learning Solutions Architect

 |

Agenda

Day 1	Day 2	Day 3
AI/ML on AWS Intro lab	Feature engineering Model evaluation	Moving to production Build
Team up Define problem	Build	Present
Write-up	Working model	Solution architecture

Advanced Sagemaker Topics
Architecture Patterns, Security and DevOps

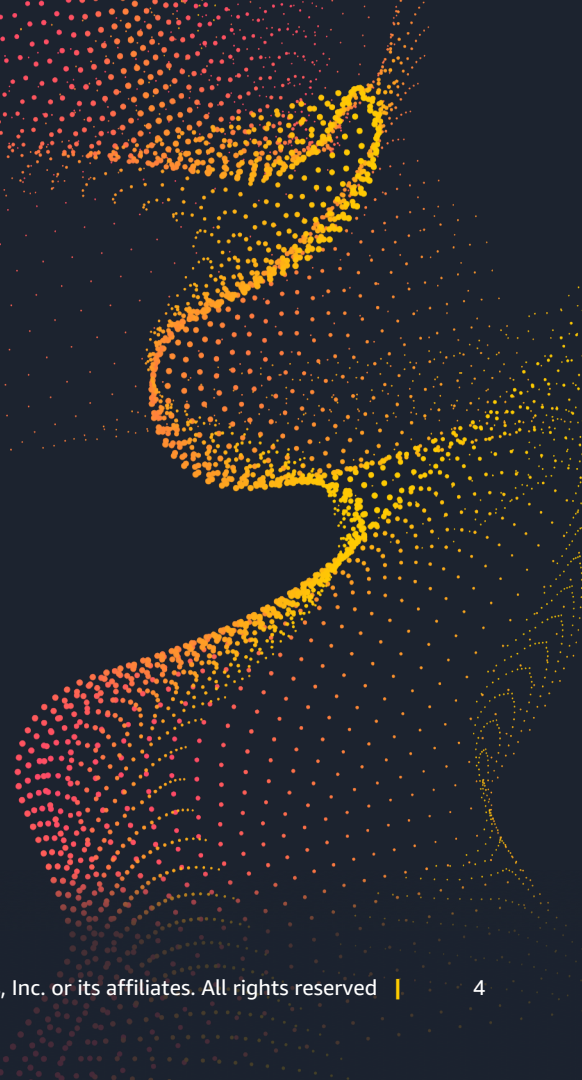
aws machine learning

Survey link

http://bit.ly/2M3QF8j

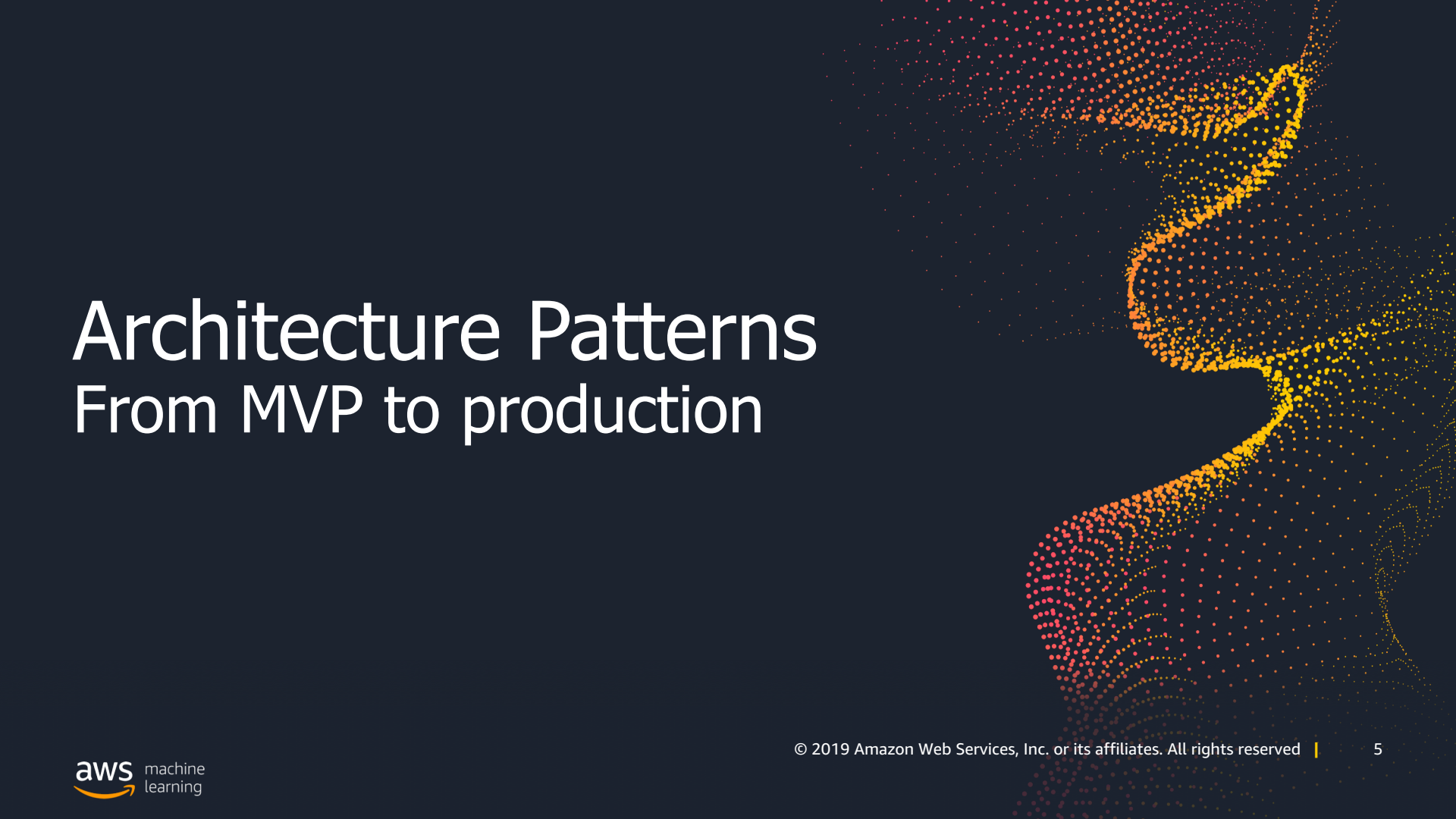
Presentations PDF

https://bit.ly/2kRXfVZ



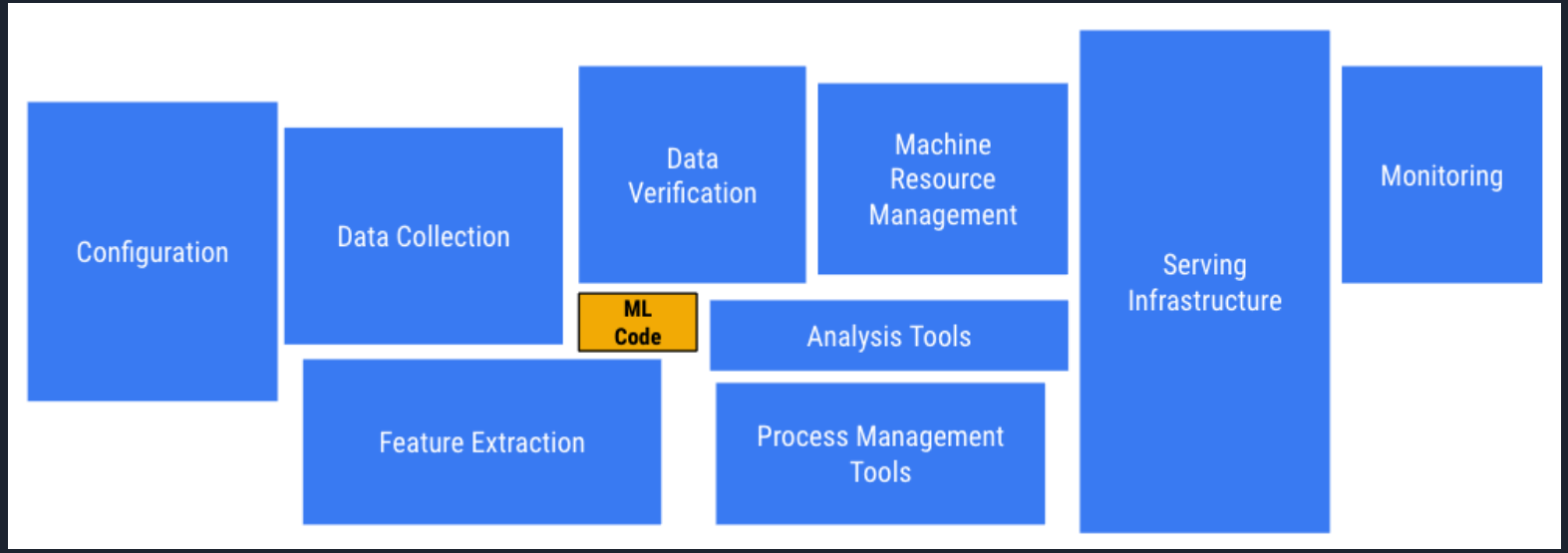
aws machine learning

Architecture Patterns
From MVP to production



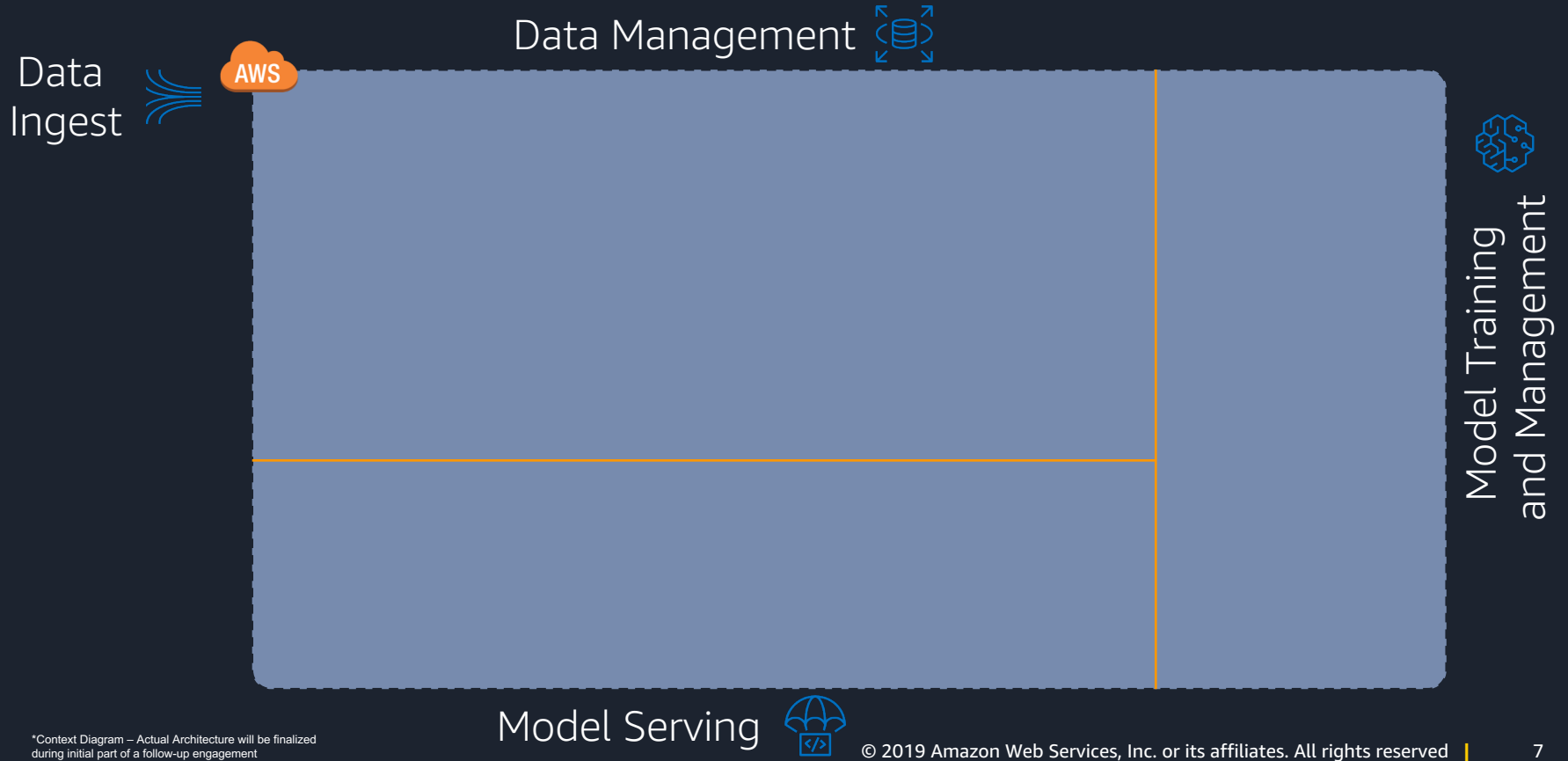
aws machine learning

Hidden technical debt of Machine Learning



aws machine learning

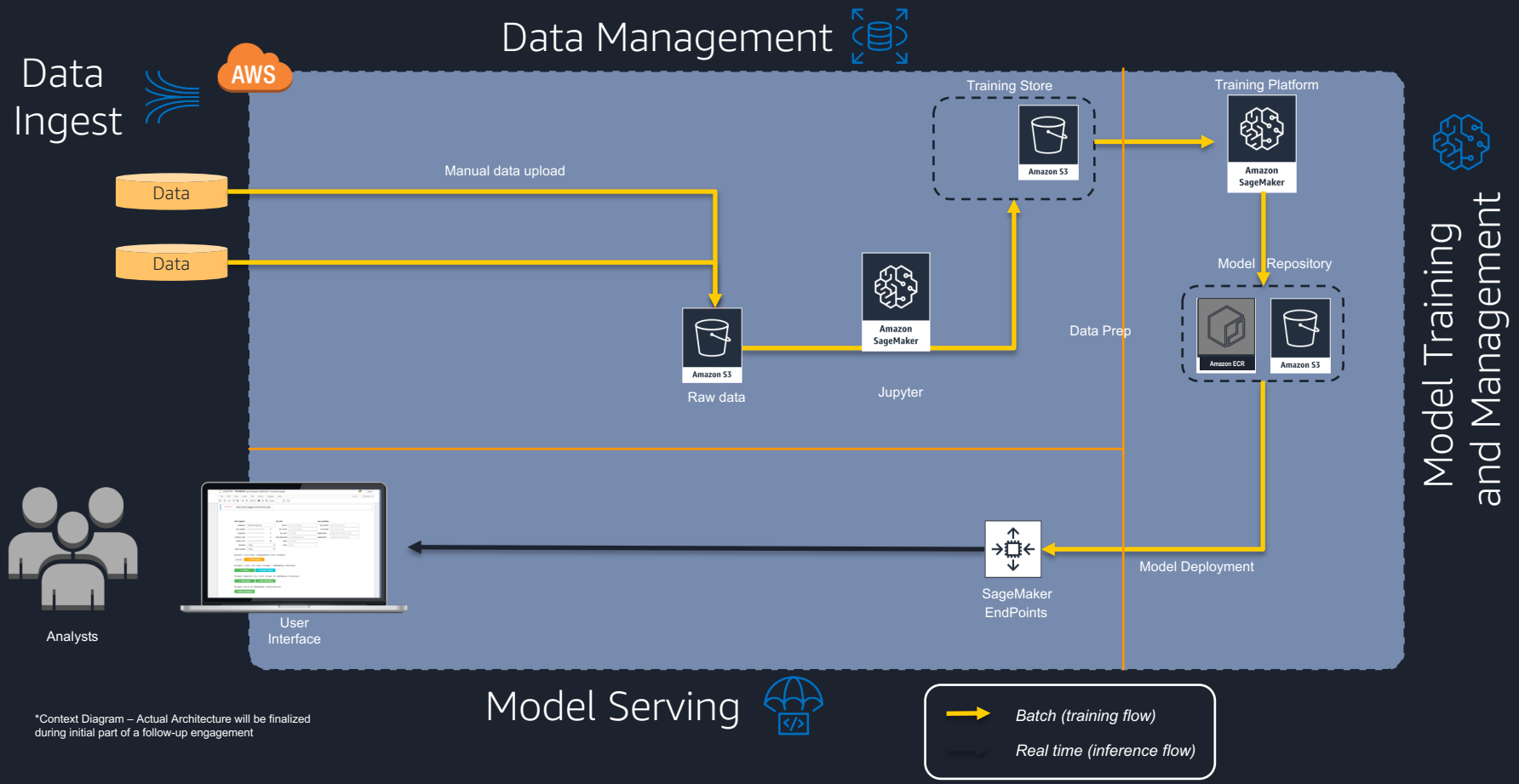
Architecture building blocks



*Context Diagram – Actual Architecture will be finalized during initial part of a follow-up engagement

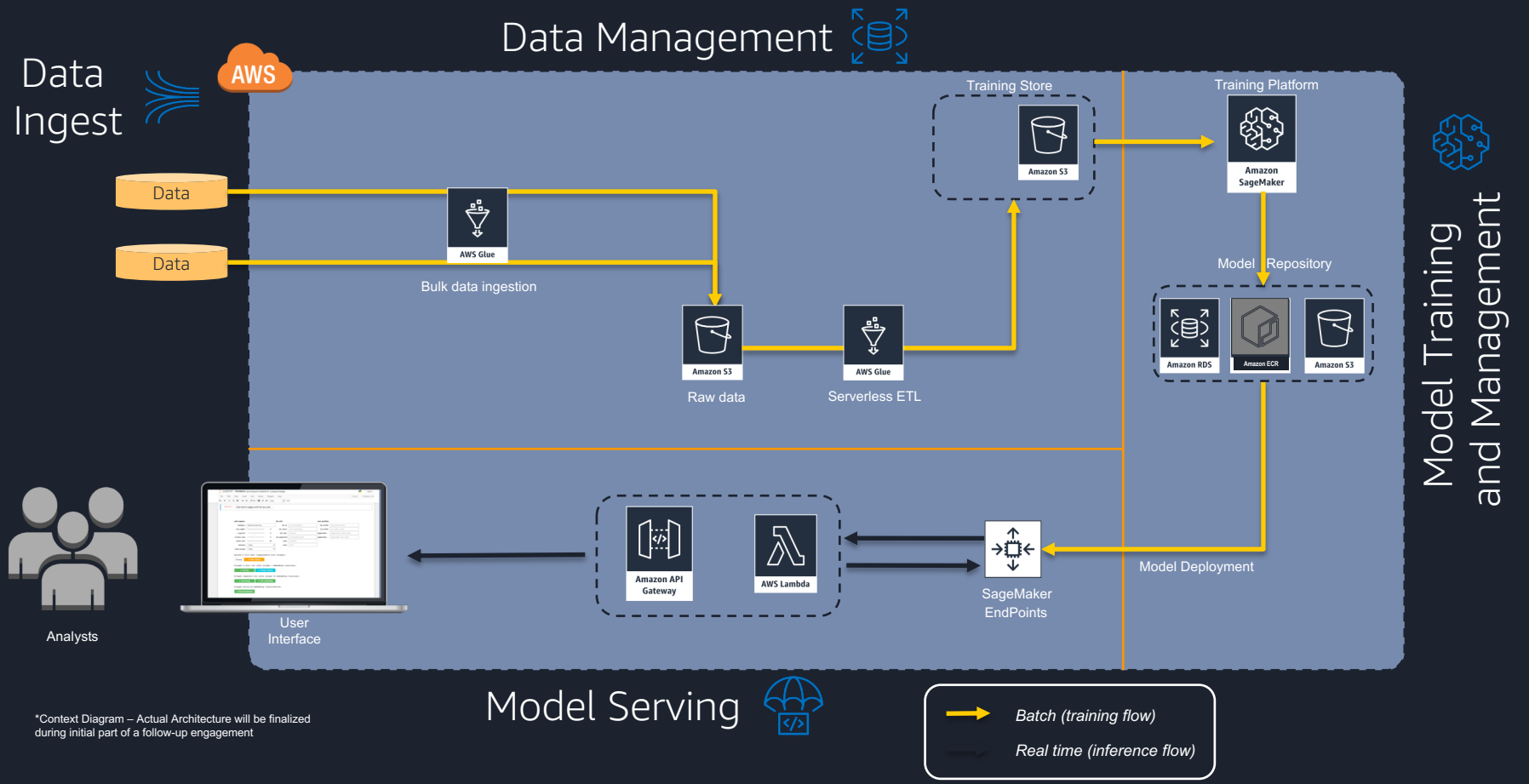
aws machine learning

POC/MVP



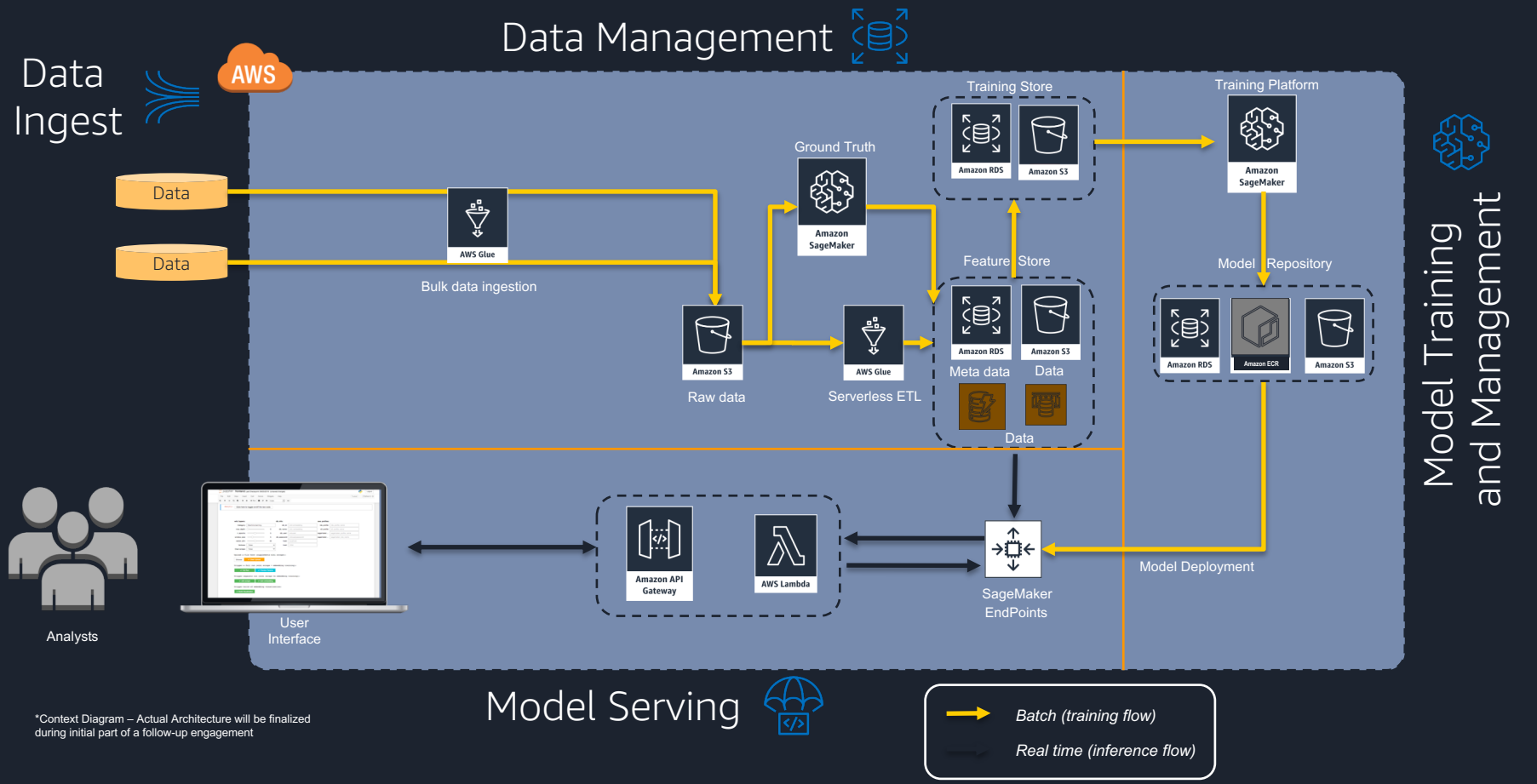
*Context Diagram – Actual Architecture will be finalized during initial part of a follow-up engagement

Model management enhancement



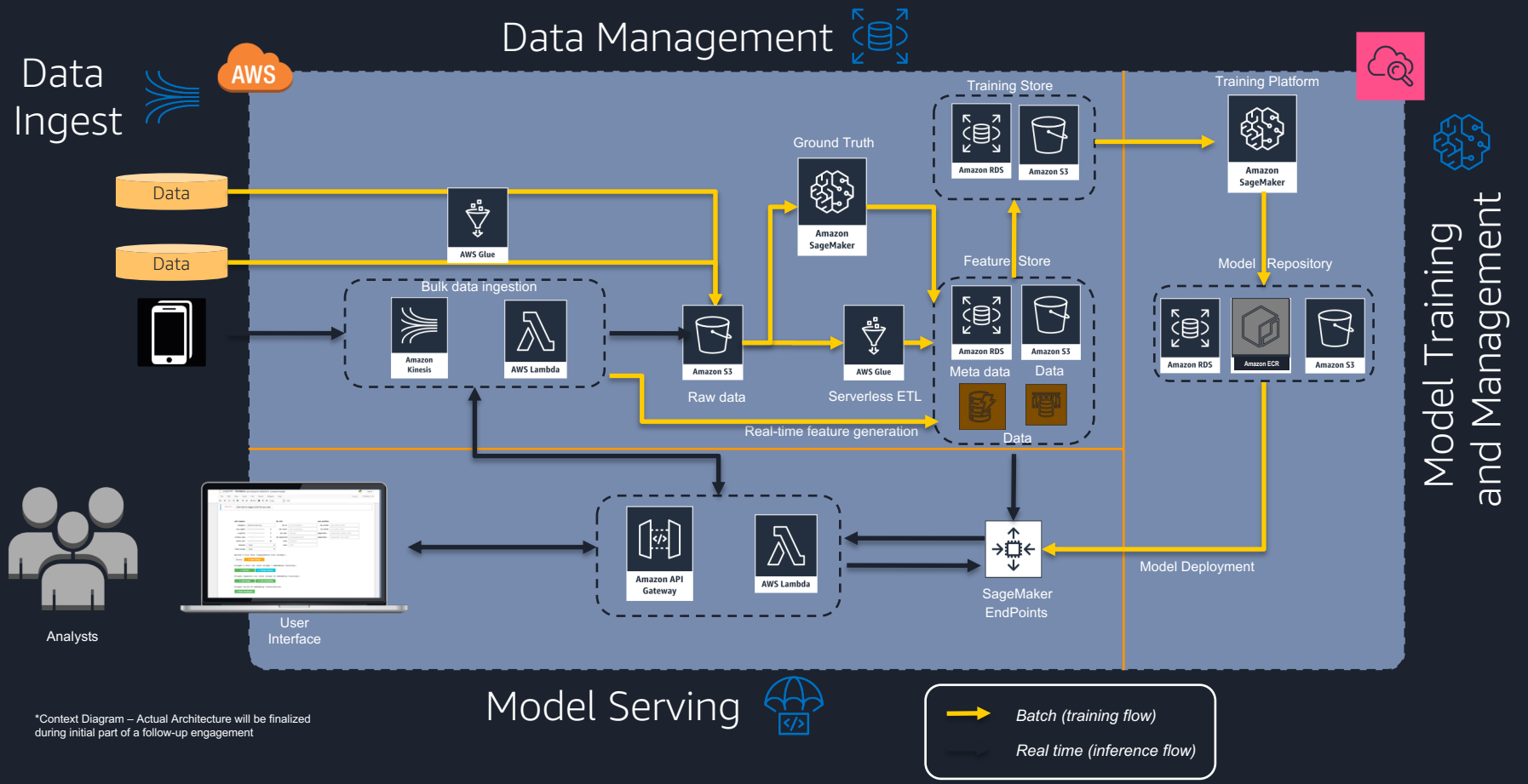
*Context Diagram – Actual Architecture will be finalized during initial part of a follow-up engagement

Feature and training store management



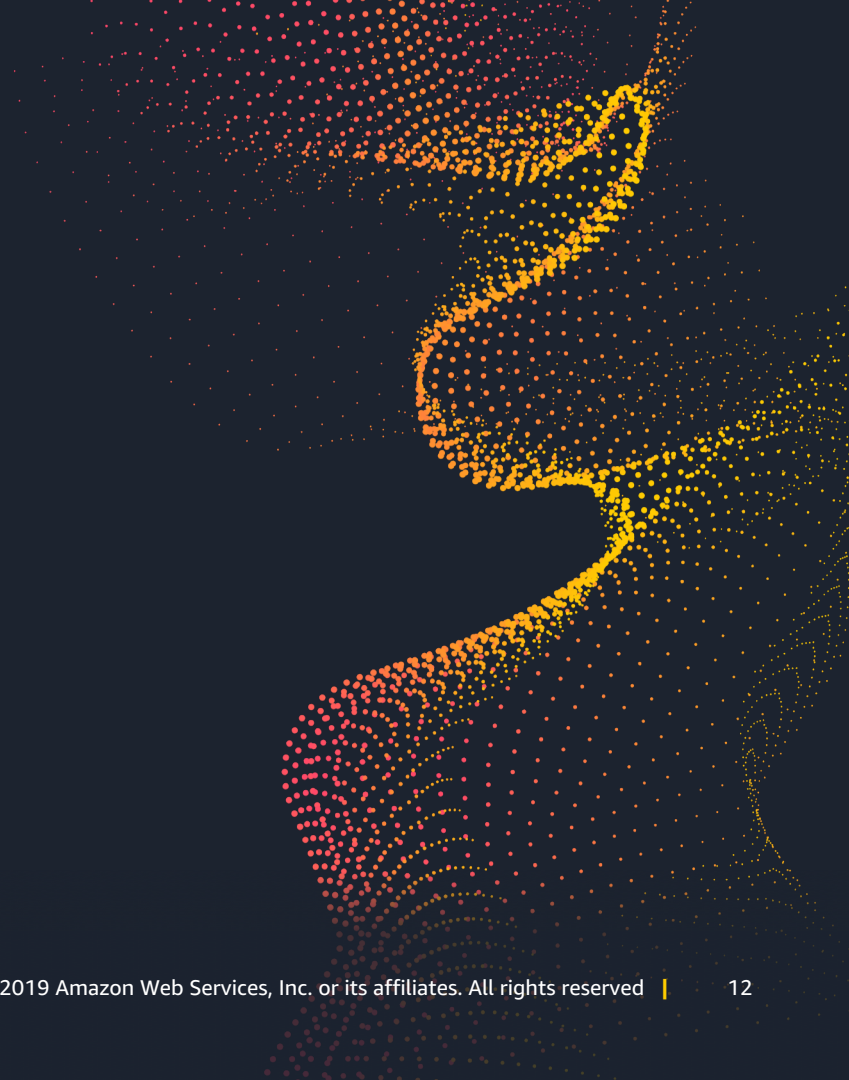
*Context Diagram – Actual Architecture will be finalized during initial part of a follow-up engagement

Realtime capability and Monitoring



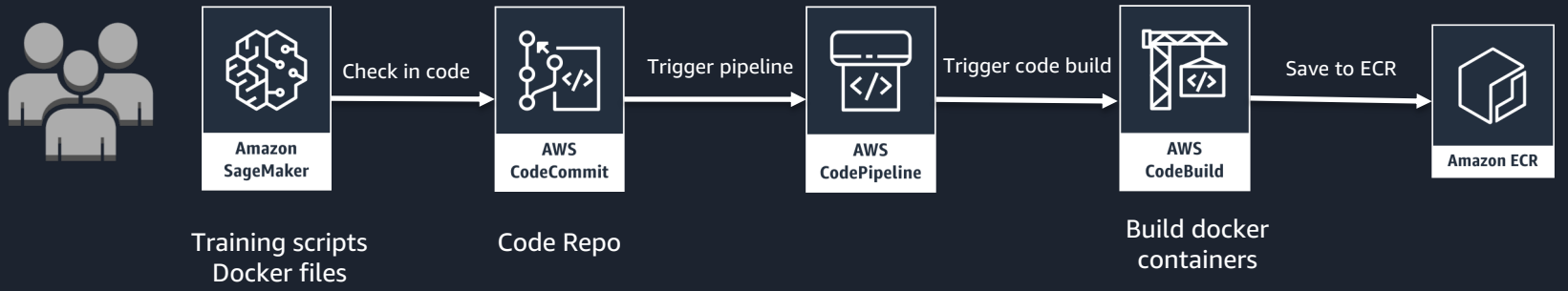
*Context Diagram – Actual Architecture will be finalized during initial part of a follow-up engagement

MLOps



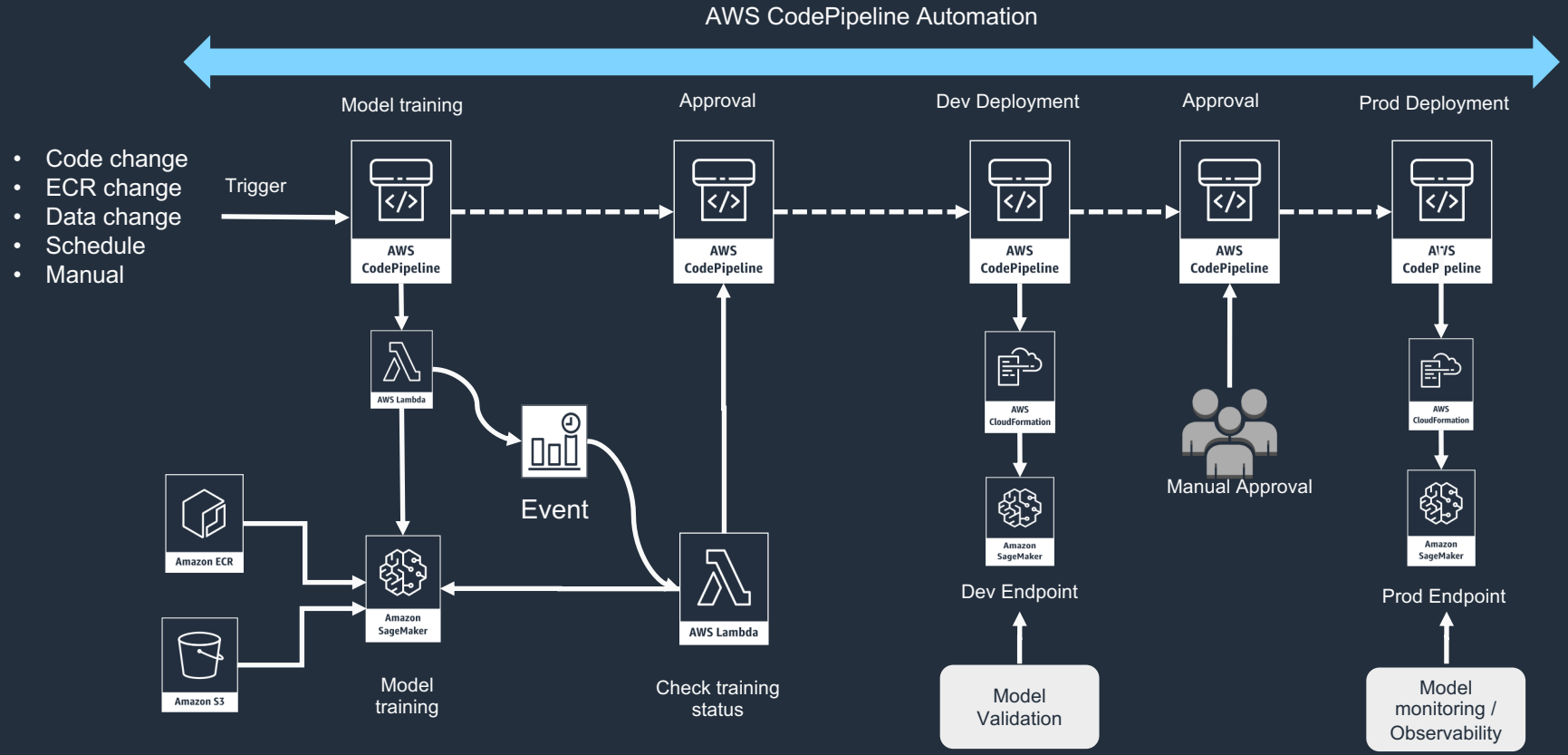
aws machine learning

ML DevOps – Custom Docker Image Build Pipeline



aws machine learning

Sample Training/Deploy Pipeline



How do I orchestrate my ML pipeline?

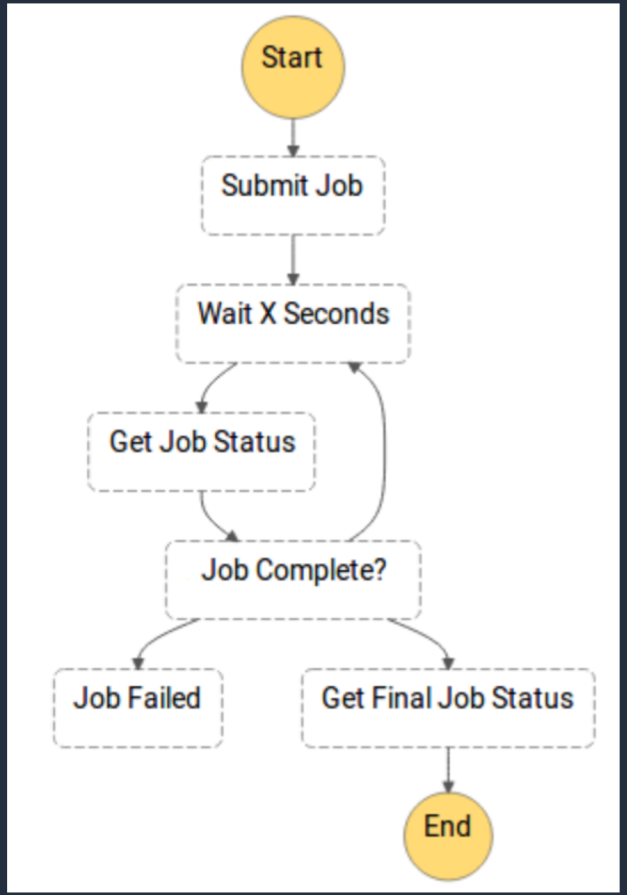
- Step functions
- Cloud Development Kit (CDK)
- Apache Airflow
- Kubeflow

Step Functions

Supported Amazon SageMaker APIs :

CreateTrainingJob

CreateTransformJob



CDK example

API gateway > Lambda > SM


```
lambdaFn = lambda_.Function(
    self,
    "callsmlambda",
    code=lambda_.InlineCode(handler_code),
    handler="index.lambda_handler",
    timeout=core.Duration.seconds(300),
    runtime=lambda_.Runtime.PYTHON_3_7,
    environment={"endpoint_name":endpoint_name, # CHANGE TO YOUR ENDPOINT NAME!!
                 "content_type":"text/csv"}
)

lambdaFn.add_to_role_policy(aws_iam.PolicyStatement(actions=['sagemaker:InvokeEndpoint',],
    resources = ['arn:aws:sagemaker:{}:{}:endpoint/{}'.format(my_region,my_acc_id,endpoint_name),]))
```

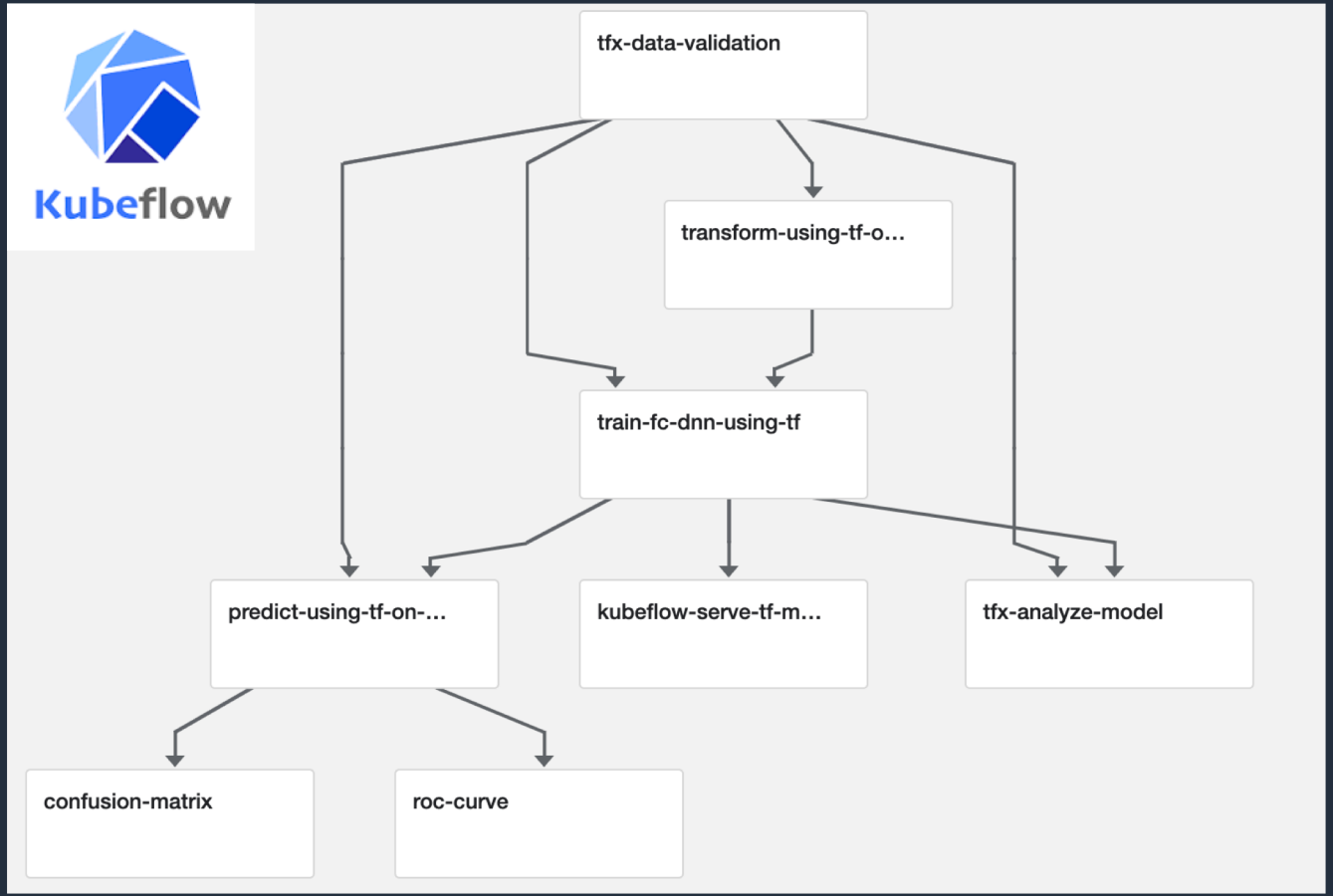
Airflow with SageMaker on AWS

Currently, the following SageMaker operators are supported:

- SageMakerTrainingOperator
- SageMakerTuningOperator
- SageMakerModelOperator
- SageMakerTransformOperator
- SageMakerEndpointConfigOperator
- SageMakerEndpointOperator

https://sagemaker.readthedocs.io/en/stable/using_workflow.html

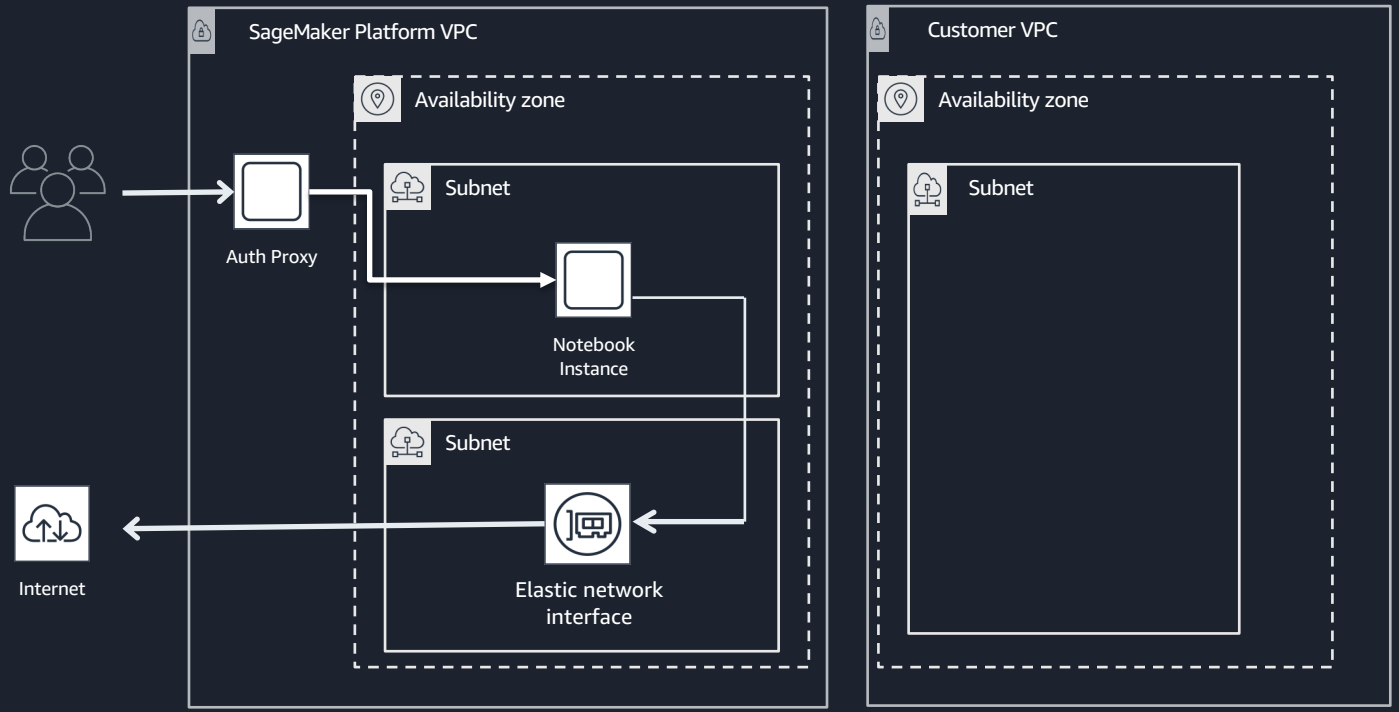
Taxi Tip Prediction Model Trainer using Tensorflow DNN



Notebook Security

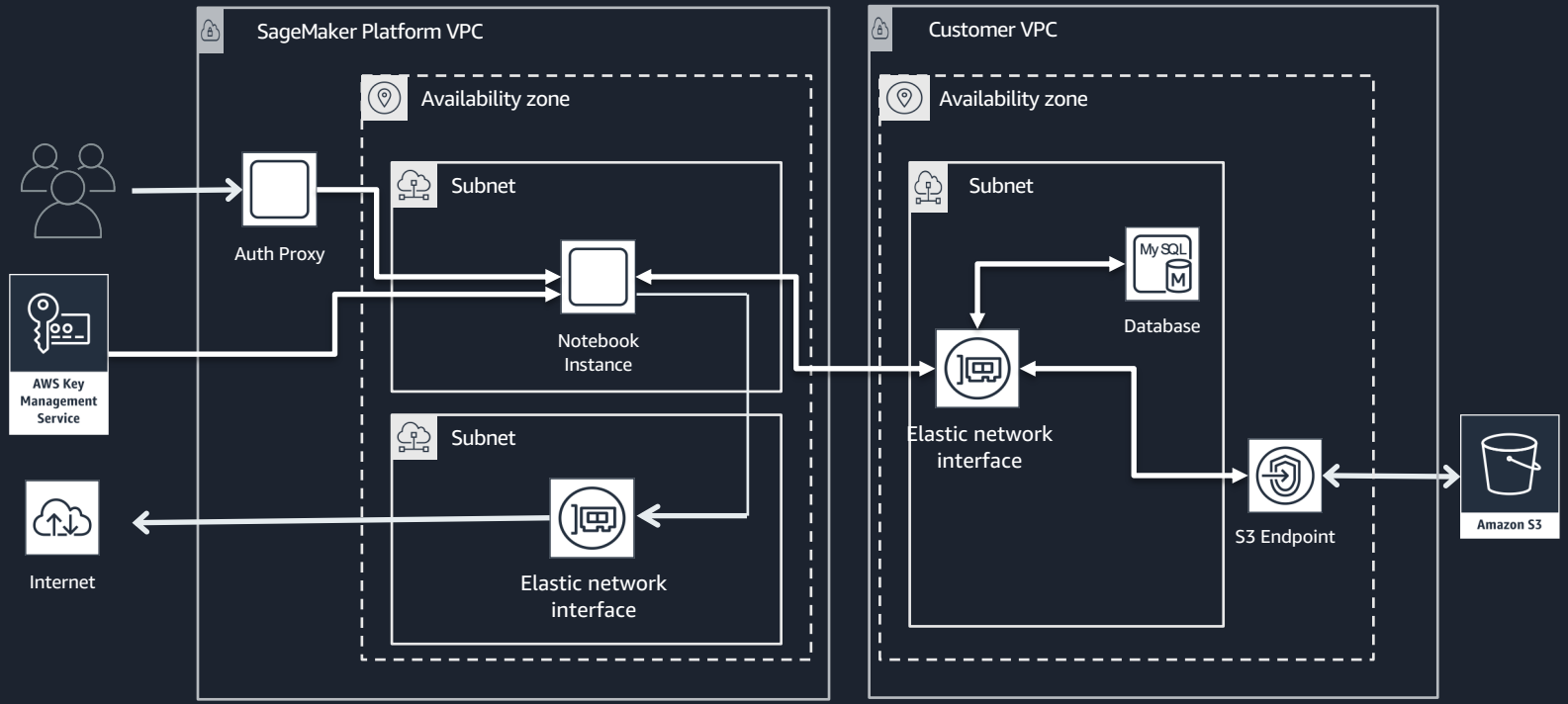
aws machine learning

Jupyter notebooks - Default



aws machine learning

Secure Jupyter notebooks + VPC



aws machine learning

Deployment Best Practices

aws machine learning

Deployment best practices

- Deploy to at least 2 availability zones, multiple regions for DR
- Use auto-scaling to accommodate traffic spikes
- Consider Lambda and API Gateway
- Consider batch inference or edge deployment
- Use Elastic Inference Accelerators to reduce inference costs
- Automate training and deployment pipeline
- Use blue/green deployment techniques
- Clearly define re-training triggers
- Enable model versioning, implement MLDevOps pipeline
- Monitor performance

Project Presentations

- Project goal, business context
- ETL strategy
- Modeling strategy
- Solution architecture
- Goal for going into production
- 20 minutes to present, can use slides

