Train Machine Learning Models using Amazon SageMaker with TensorFlow

Ahmad R Khan

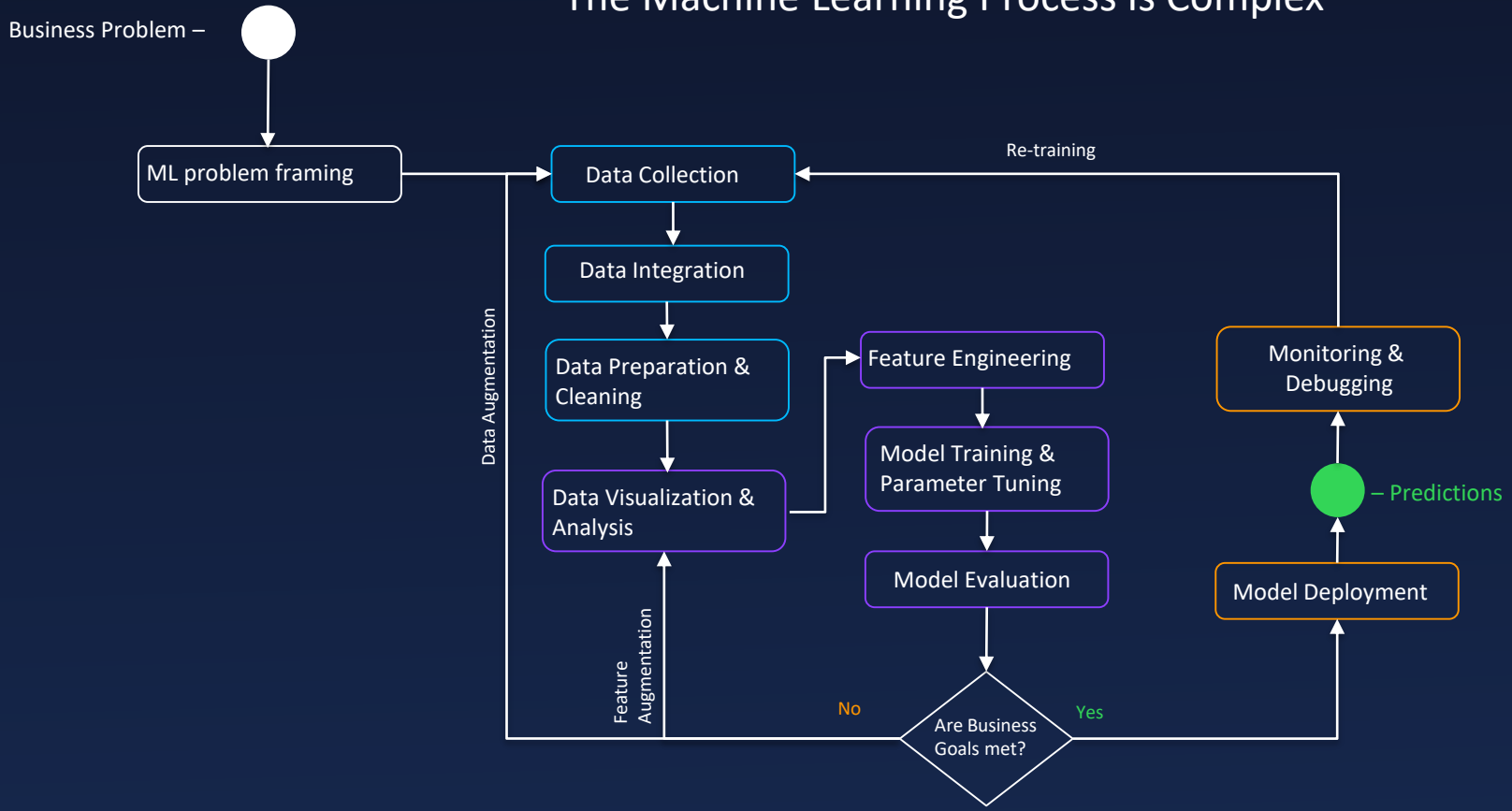
Solutions Architect, AWS

Questions We'll Answer In This Session:

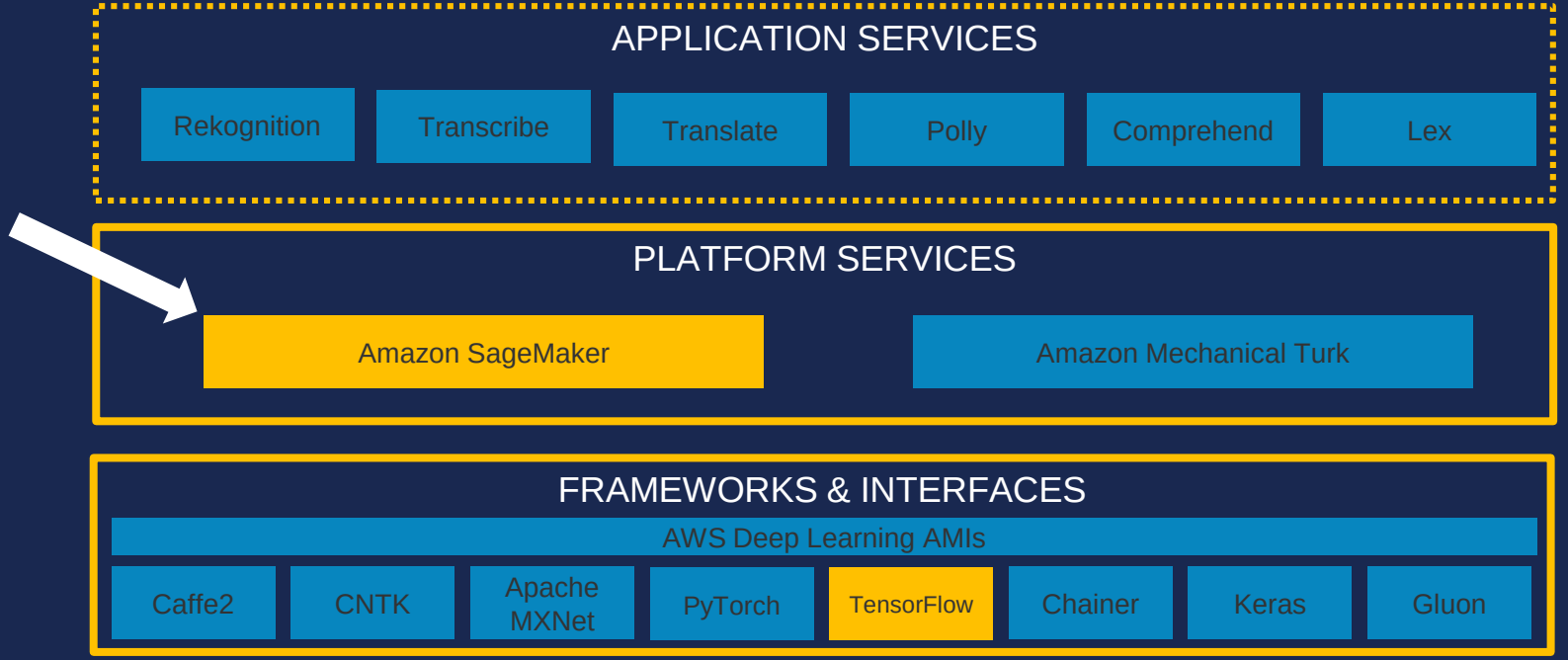
1. Amazon SageMaker: what is it? how does it work?
2. Why run TensorFlow on SageMaker?
3. How to train a TensorFlow model using SageMaker? (Demo)
4. How to host a trained TensorFlow model on SageMaker to provide scalable inferencing service? (Demo)
5. How to get started?

1. Amazon SageMaker:
    What is it?
    How does it work?

The Machine Learning Process Is Complex



The Amazon Machine Learning Stack



EDUCATION

AWS DeepLens

Amazon SageMaker

Amazon SageMaker is a fully-managed platform that enables developers and data scientists to quickly and easily build, train, and deploy machine learning models at any scale.

Amazon SageMaker removes all the barriers that typically slow down developers who want to use machine learning.

aws

Amazon SageMaker

Build

Pre-built notebook instances

Highly-optimized machine learning algorithms

Train

One-click training for ML, DL, and custom algorithms

Easier training with hyperparameter optimization

TensorFlow

mxnet

PYTORCH

GLUON

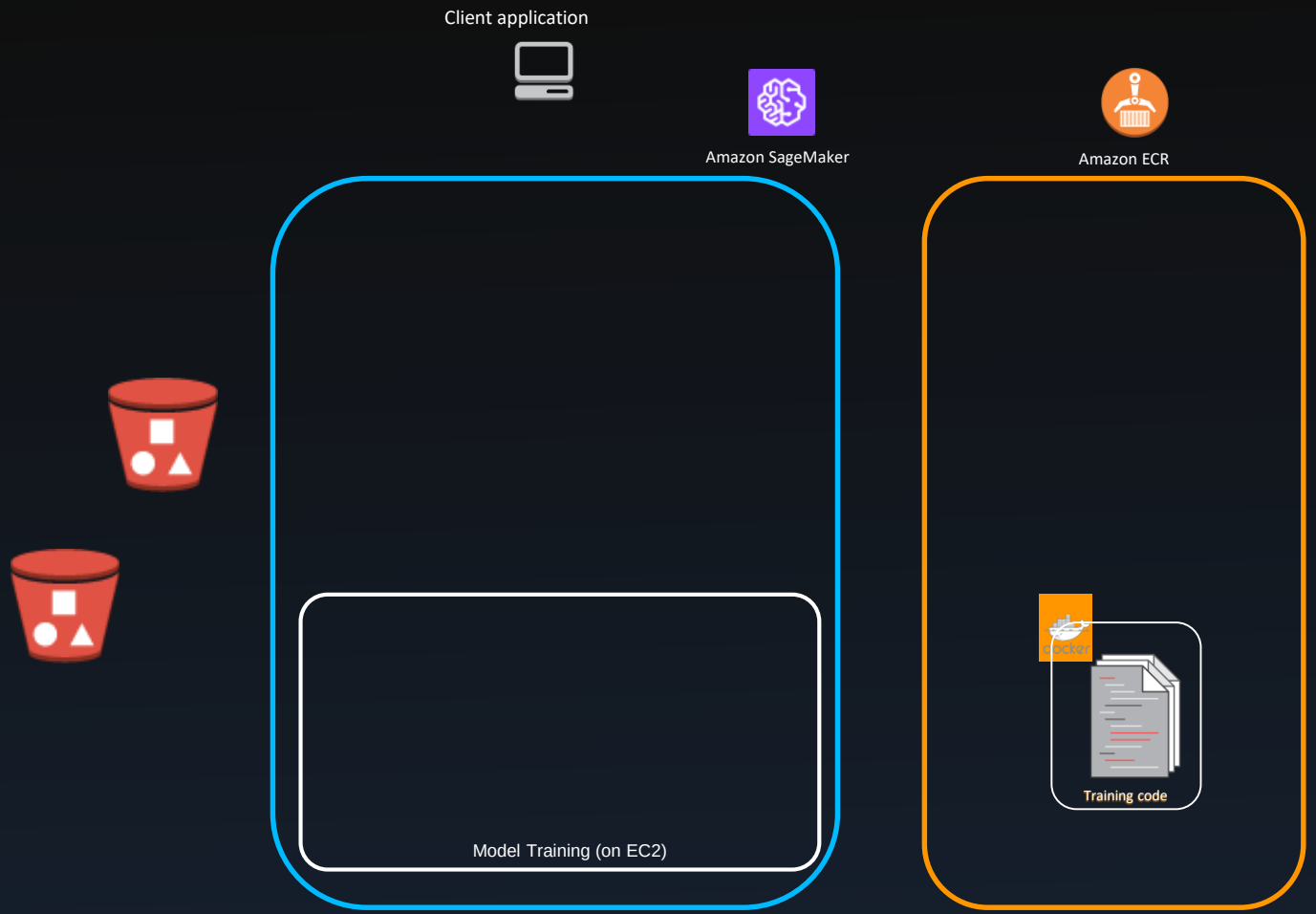
scikit learn

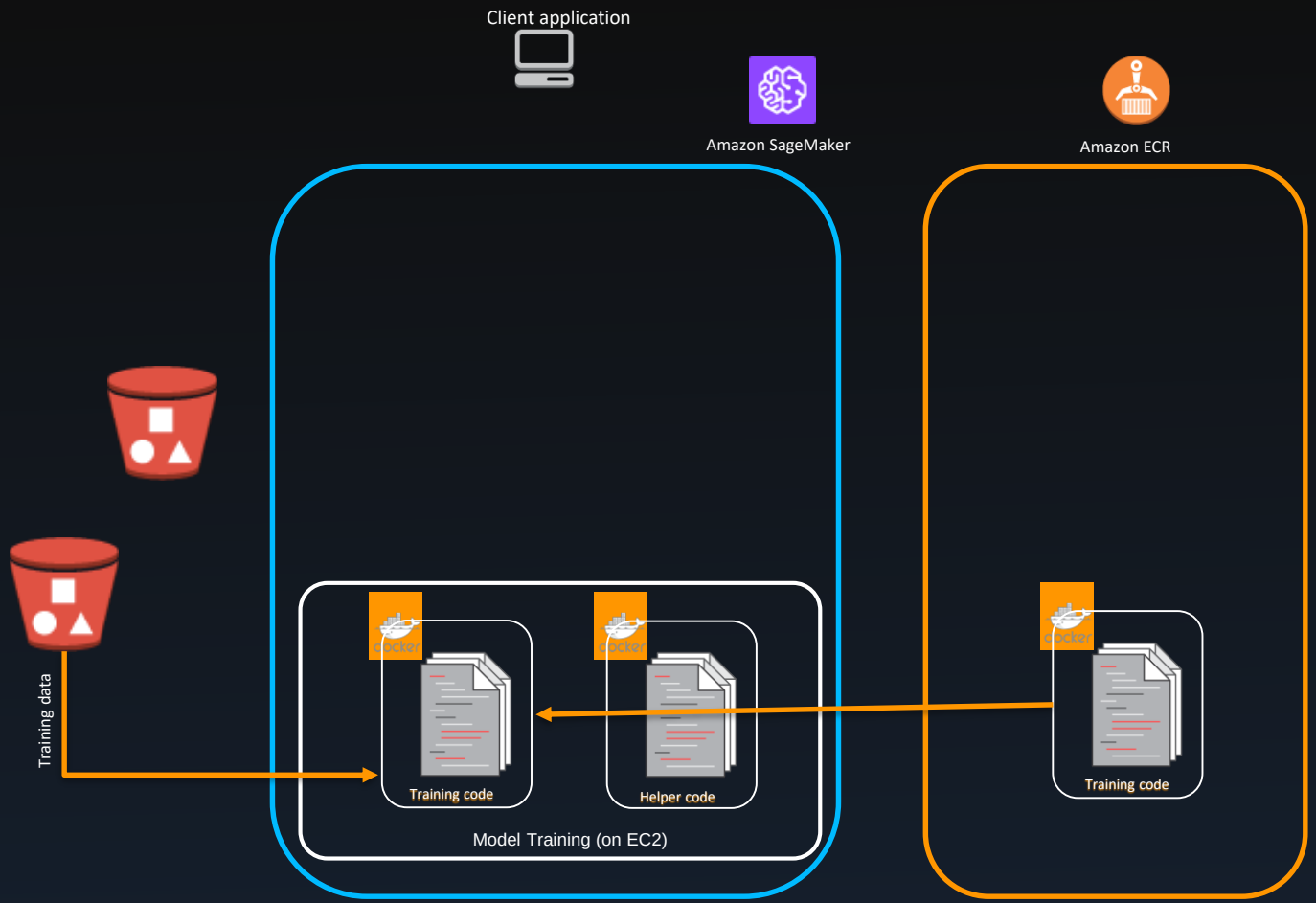
Deploy

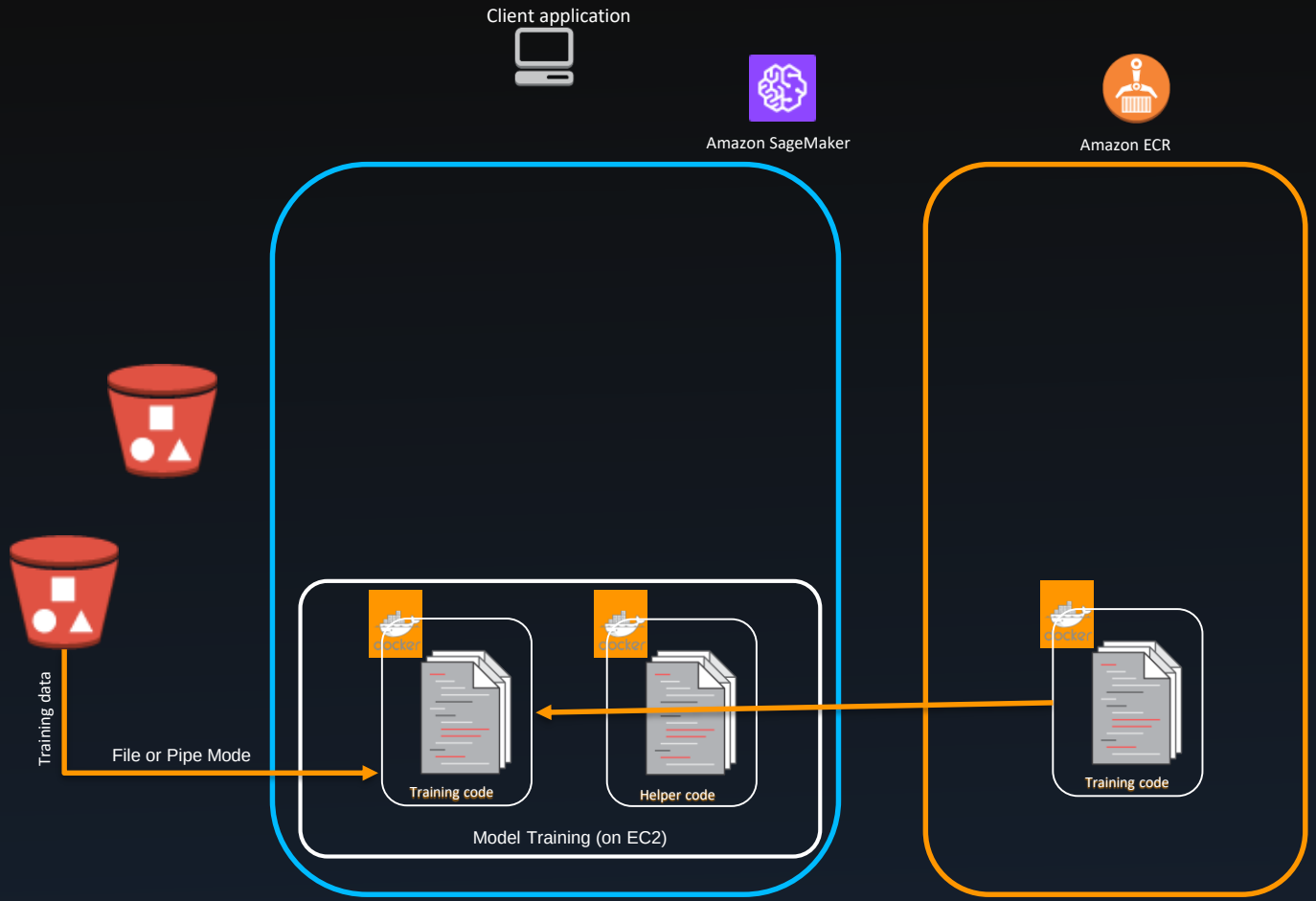
Deployment without engineering effort

Fully-managed hosting at scale

aws







SageMaker Data Input Modes

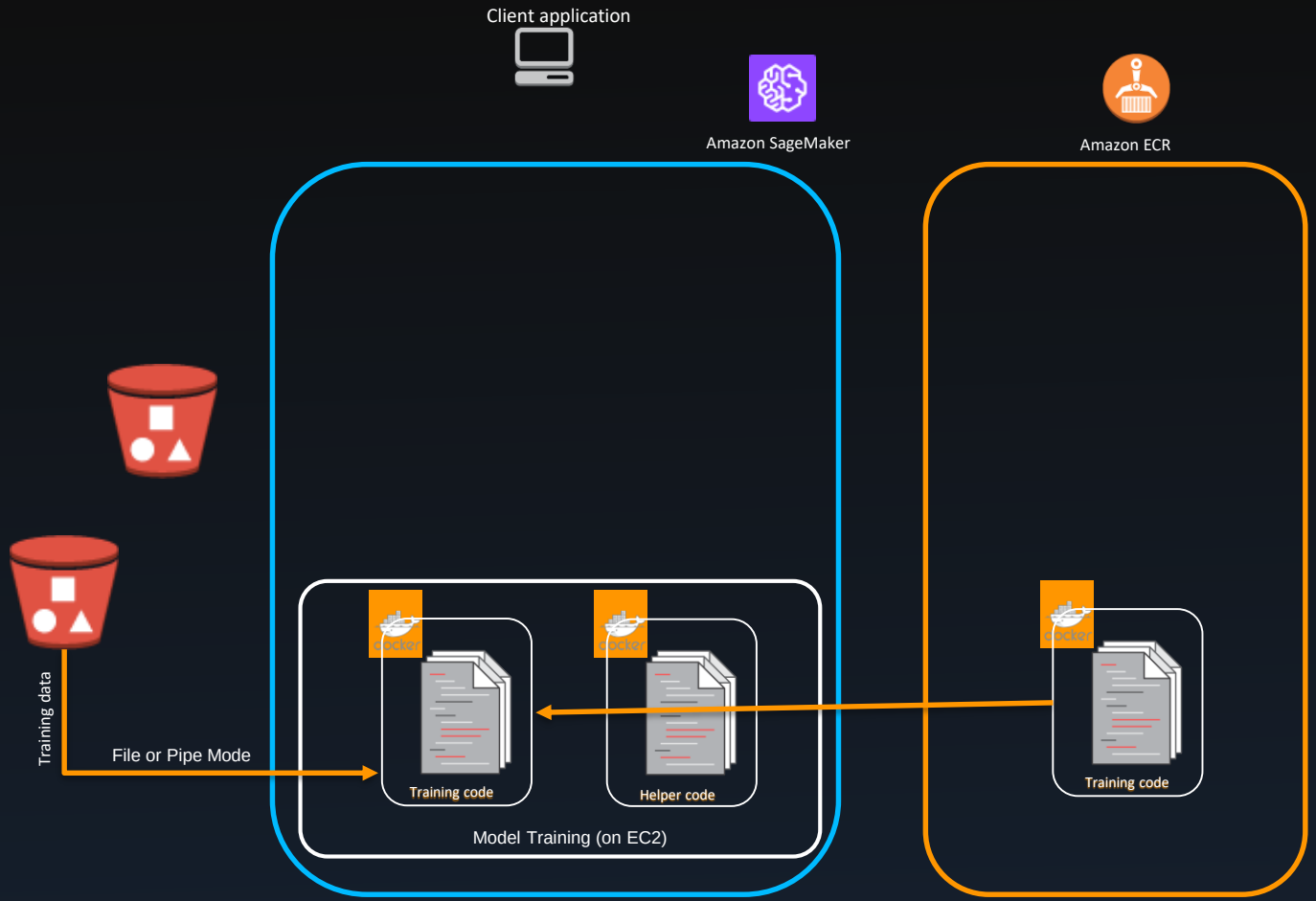
File Mode

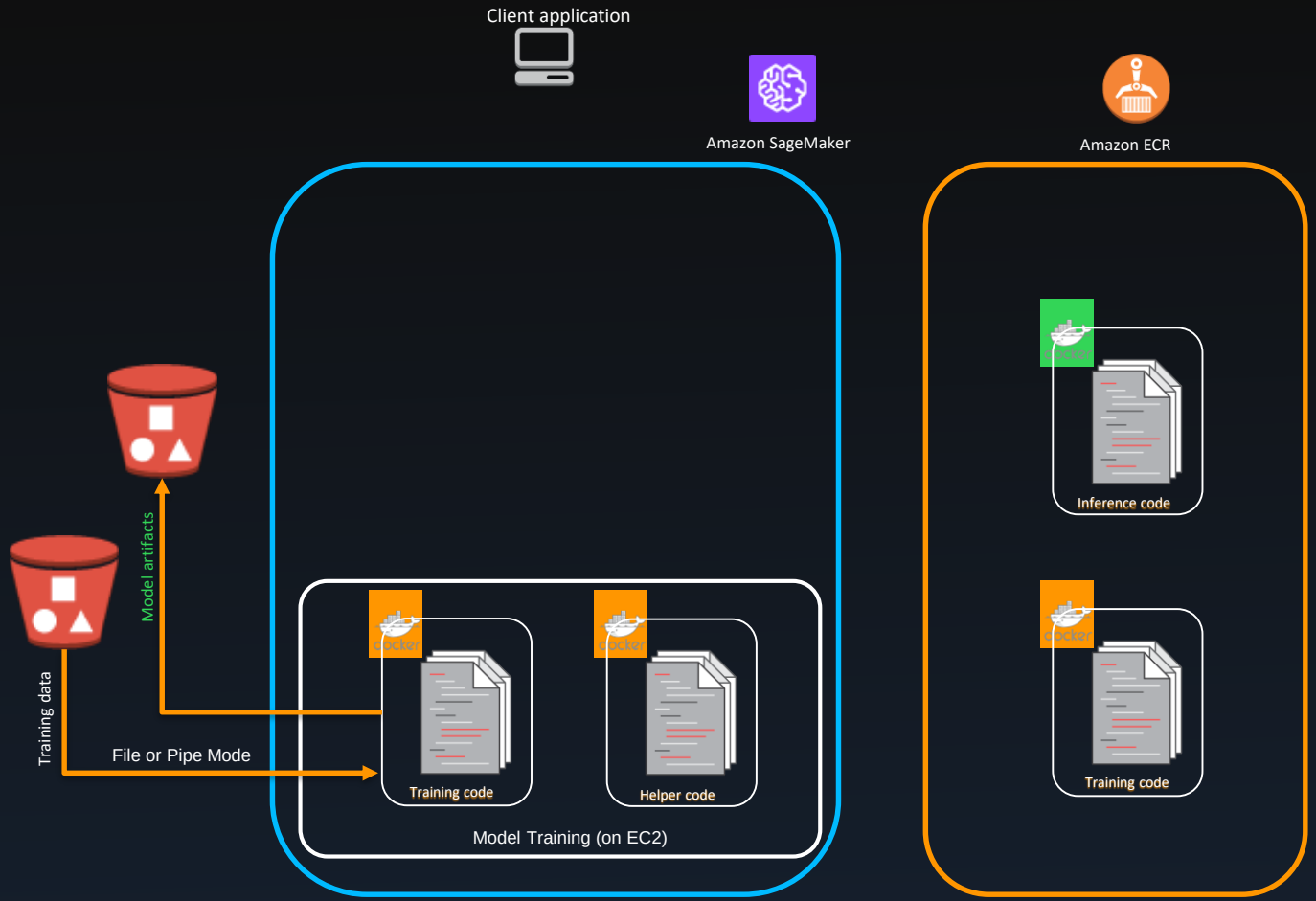
- Copies all data files from S3 to training instance volume
- Works with any supported data format
- Needs enough disk space to store entire dataset

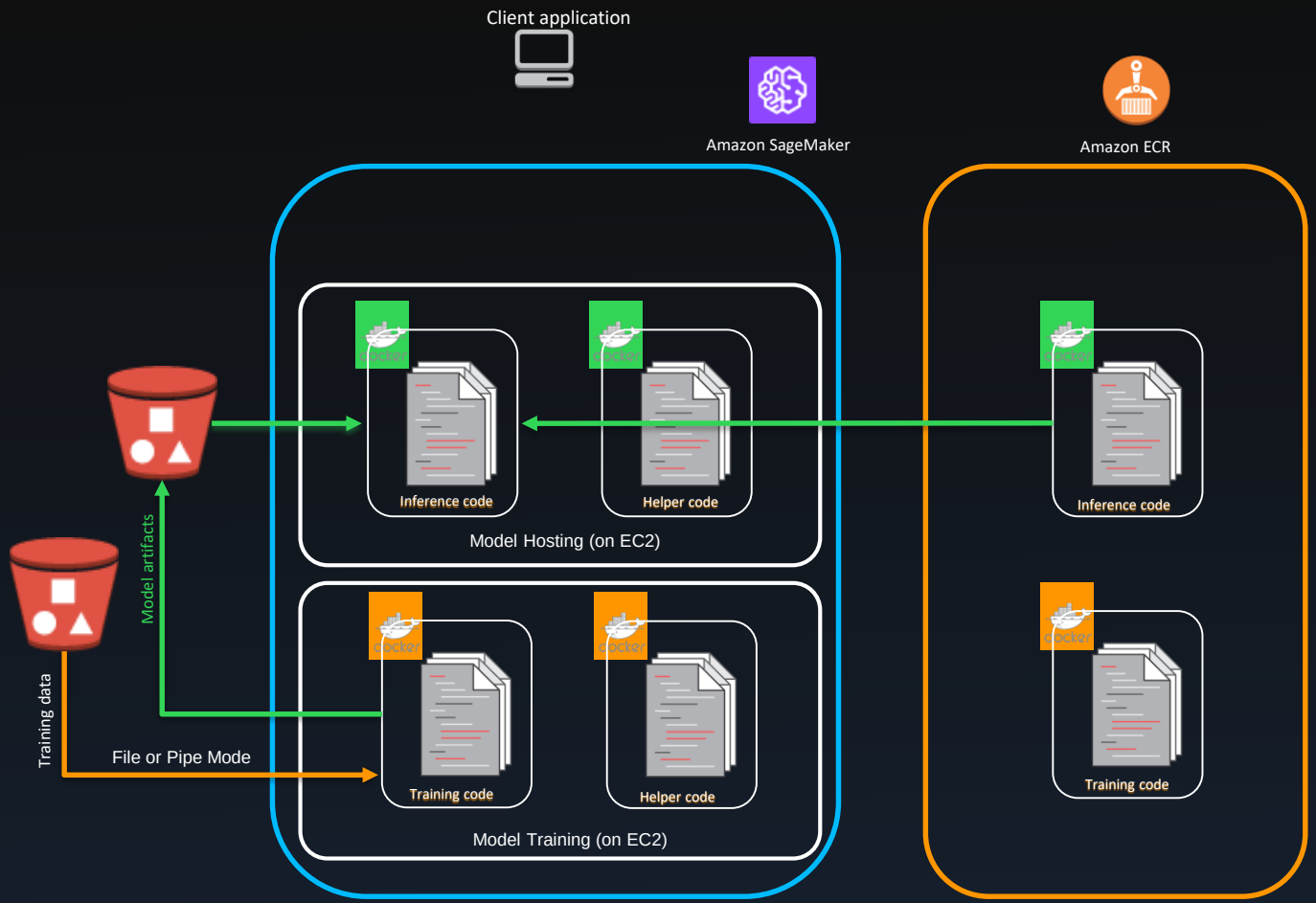
Pipe Mode

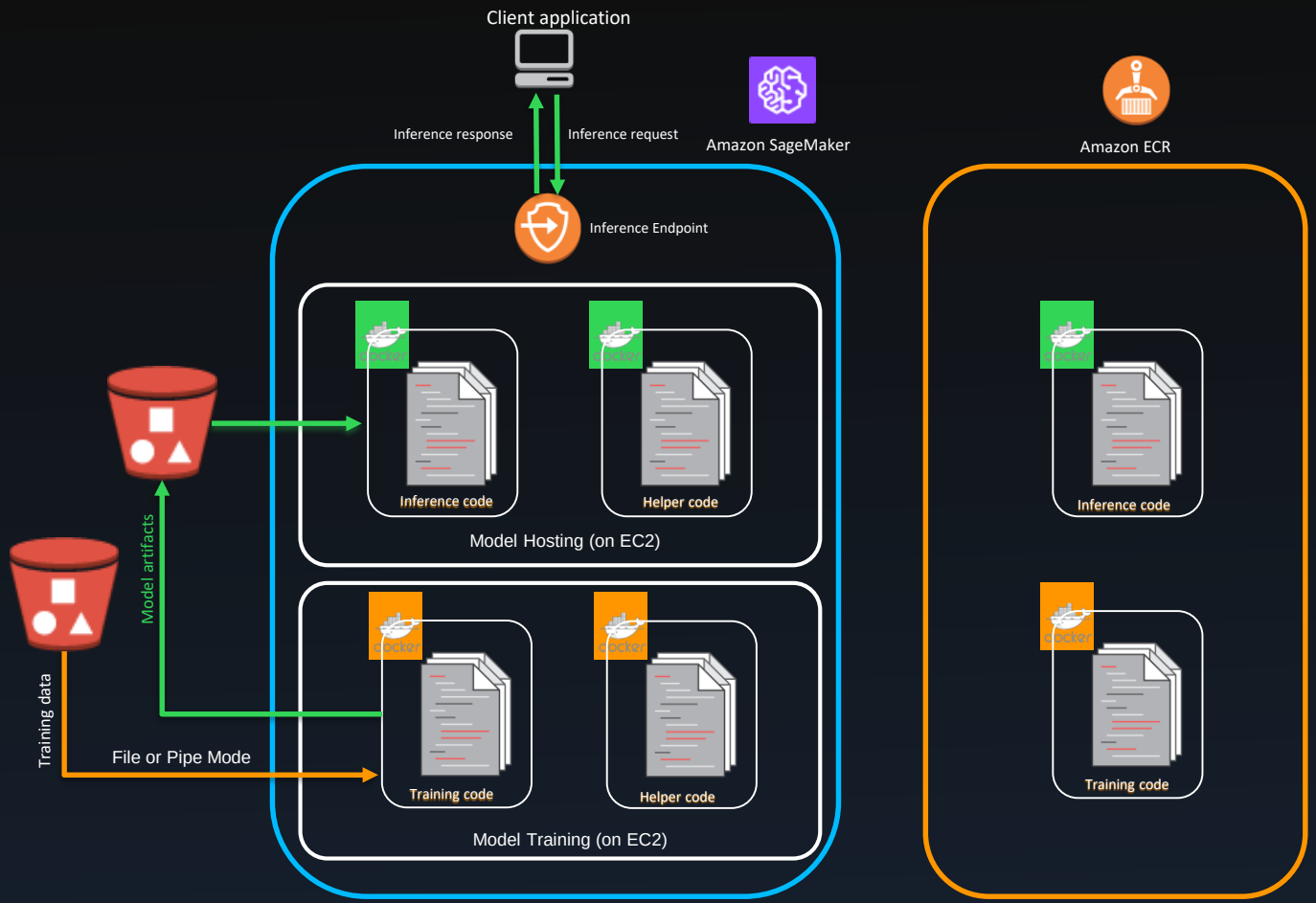
- Streams data directly from S3
- Faster start times for training jobs & better throughput
- Needs disk space only to store model artifacts
- Needs data to be in protobuf recordIO format

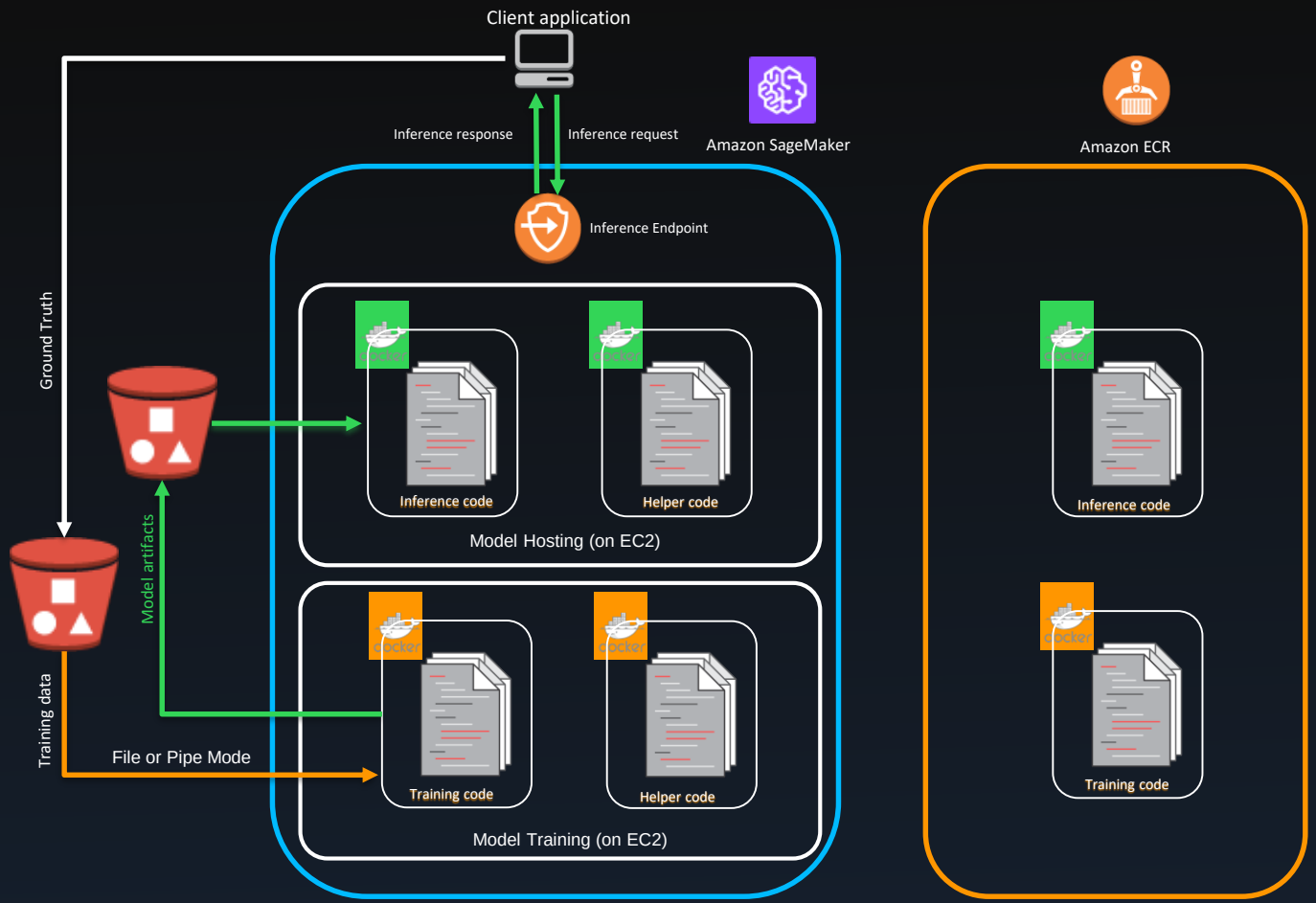
aws





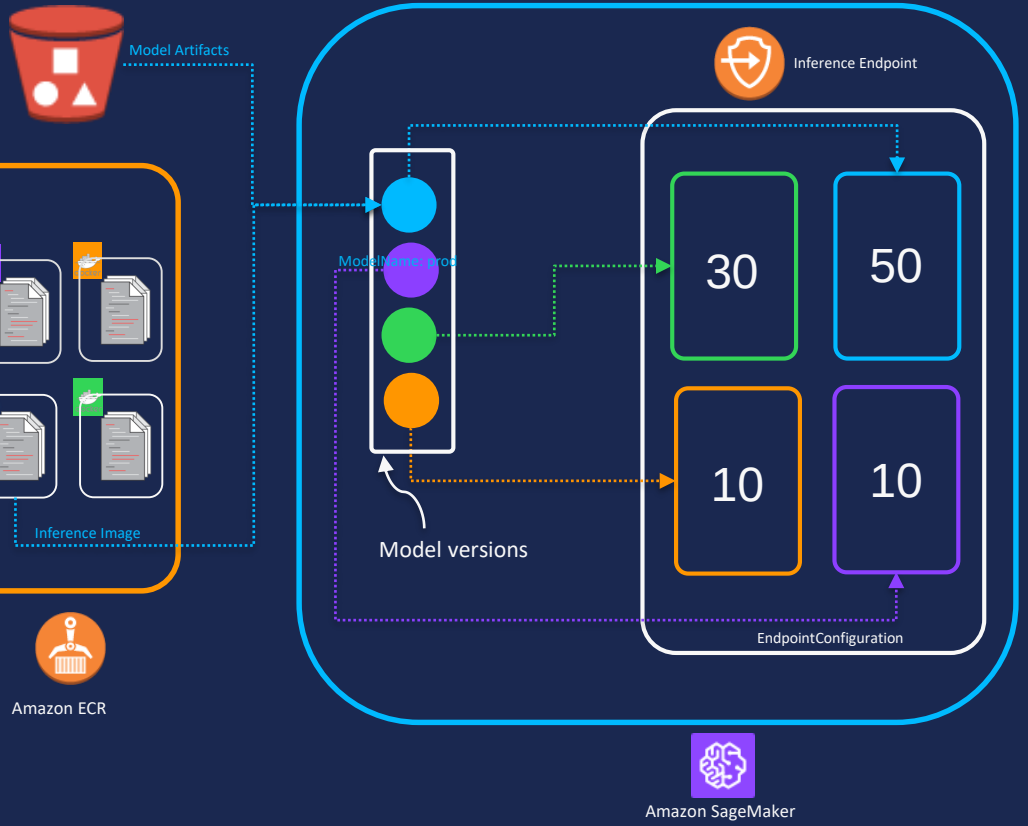






Easy Model Deployment to Amazon SageMaker – with Split Testing

ML Hosting Service



Versions of the same inference code saved in inference containers. **Prod** is the primary one, 50% of the traffic must be served there!

Create an Endpoint from one `EndpointConfiguration`

Create versions of a Model

Create weighted ProductionVariants

Create `EndpointConfiguration` from one or more `ProductionVariant(s)`

One-Click!

2. Why run TensorFlow on SageMaker?

Amazon SageMaker

Speed & agility in all phases

Training speed

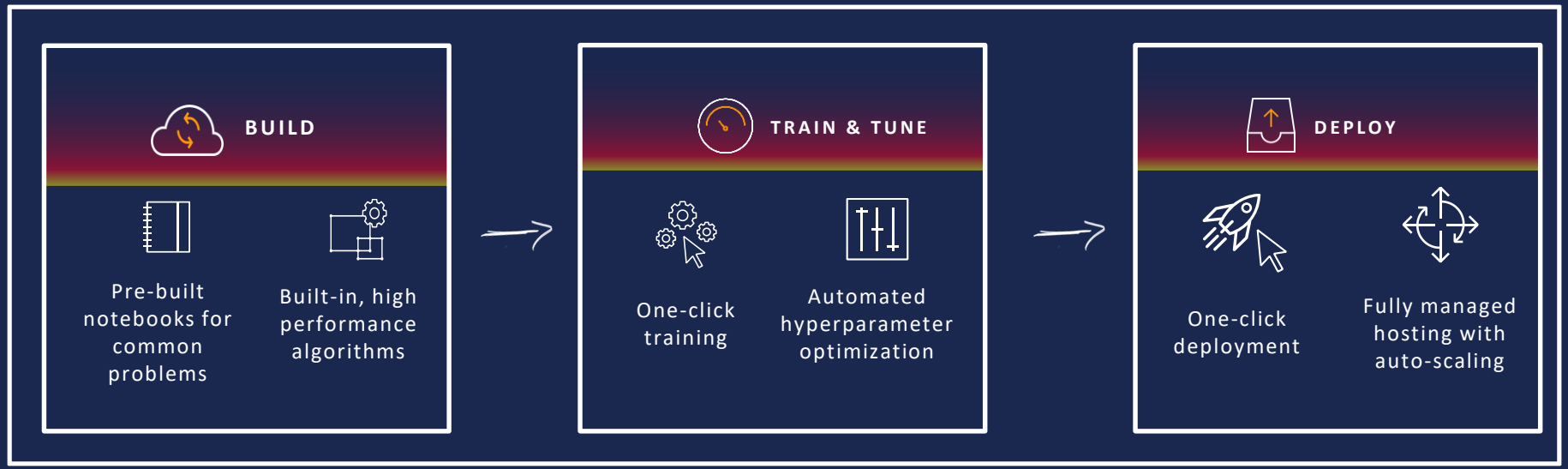
Inference speed

Remove undifferentiated heavy lifting of ML

Iterate, iterate, iterate!

Amazon SageMaker

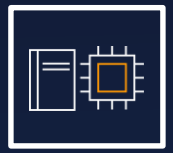
Build, train, tune, and host your own models



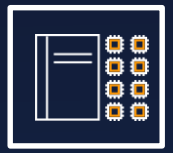
Amazon SageMaker Training Service

Enabling experimentation speed

Train with local notebooks



Train on notebook instances



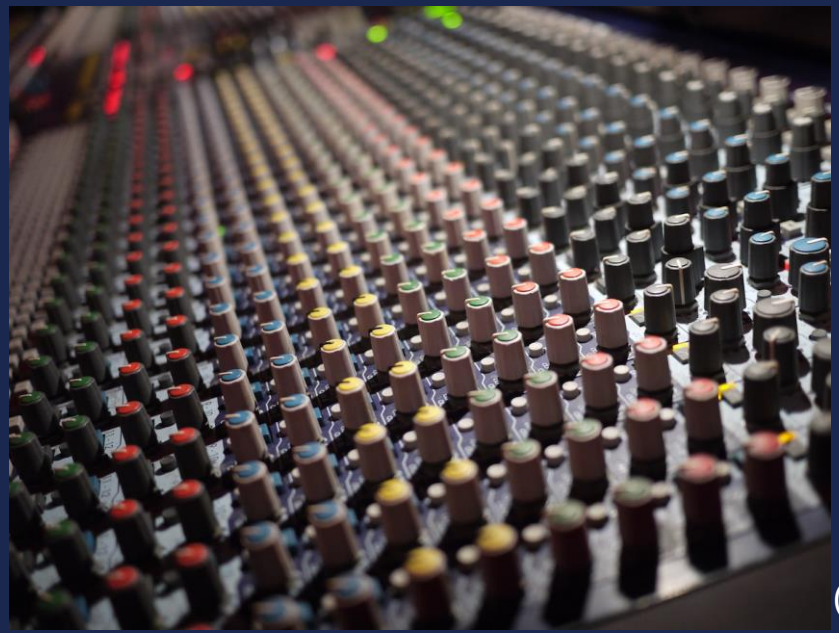
PetaFLOP training on p3.16xl

Go distributed with one line of code

Same containers

Automatic Model Tuning

Adjusting algorithm parameters to arrive at the best model



Decision Trees	Neural Networks	
Tree depth	Number of layers	
Max leaf nodes	Hidden layer width	
Gamma	Learning rate	…
Eta	Embedding dimensions	
Lambda		
Alpha	Dropout	
…	…	

"Hyperparameters"

(algorithm parameters that significantly affect model quality)

Tuning Strategy: Customized Bayesian Optimization

aws

Amazon SageMaker Model Hosting Service

- Deploys your model for inferencing as an API endpoint
  - Requires previously-trained model
- Created via web console, or via one line of code
  - In Python: sagemaker.deploy() method
- Manages the infrastructure on your behalf:
  - Amazon EC2 instances
  - docker containers
  - auto-scaling option

aws

3. How to Train a TensorFlow model using SageMaker? (Demo)

Demo:

Training *Iris* DNN Model on SageMaker Using TensorFlow Estimator (`tf.estimator`)

aws

4. How to host a trained TensorFlow model on SageMaker to provide scalable inferencing service? (Demo)

{ Return to Demo:

Hosting *Iris* DNN Model on SageMaker
for Inferencing Requests

}

aws

Summary

1. Amazon SageMaker supports the complete TensorFlow ML/DL life-cycle with speed and agility
    - Jupyter notebooks provided
    - built-in cloud-scale ML algorithms
    - fully-managed and scalable training service (incl. automated model tuning)
    - model hosting with auto-scaling
    - Full life cycle can be scripted for CI/CD

2. TensorFlow code can be brought to SageMaker either as scripts or within Docker containers.

3. We demo'd a complete example of TensorFlow coding, training, and model hosting.

5. How to get started?

Getting Started Running TensorFlow on Amazon SageMaker:

- Try Amazon SageMaker for *free* using the Free Tier (https://amzn.to/2JfdiZ0)
- Read the abundant documentation online: (https://amzn.to/2KRD06y)
- Try the many TensorFlow sample notebooks included with SageMaker (code also available on Github: http://bit.ly/2KXLPMc)
- AWS ML Blog Posts about SageMaker (https://amzn.to/2KRxAZn)
- Consult SageMaker TensorFlow Container repo on Github (http://bit.ly/2KWpDSw)

aws

Thank You!

Questions?