

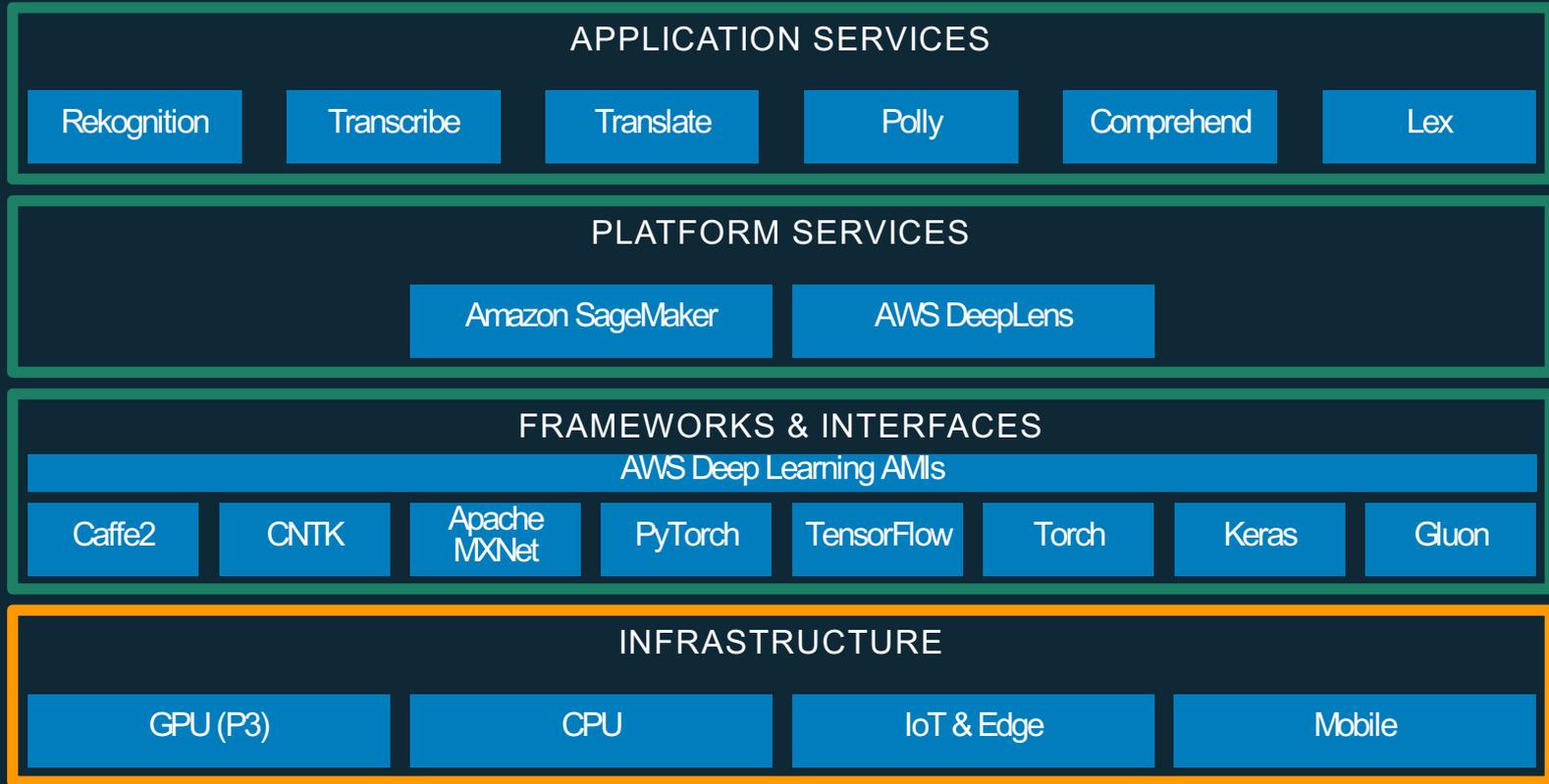


Amazon SageMaker

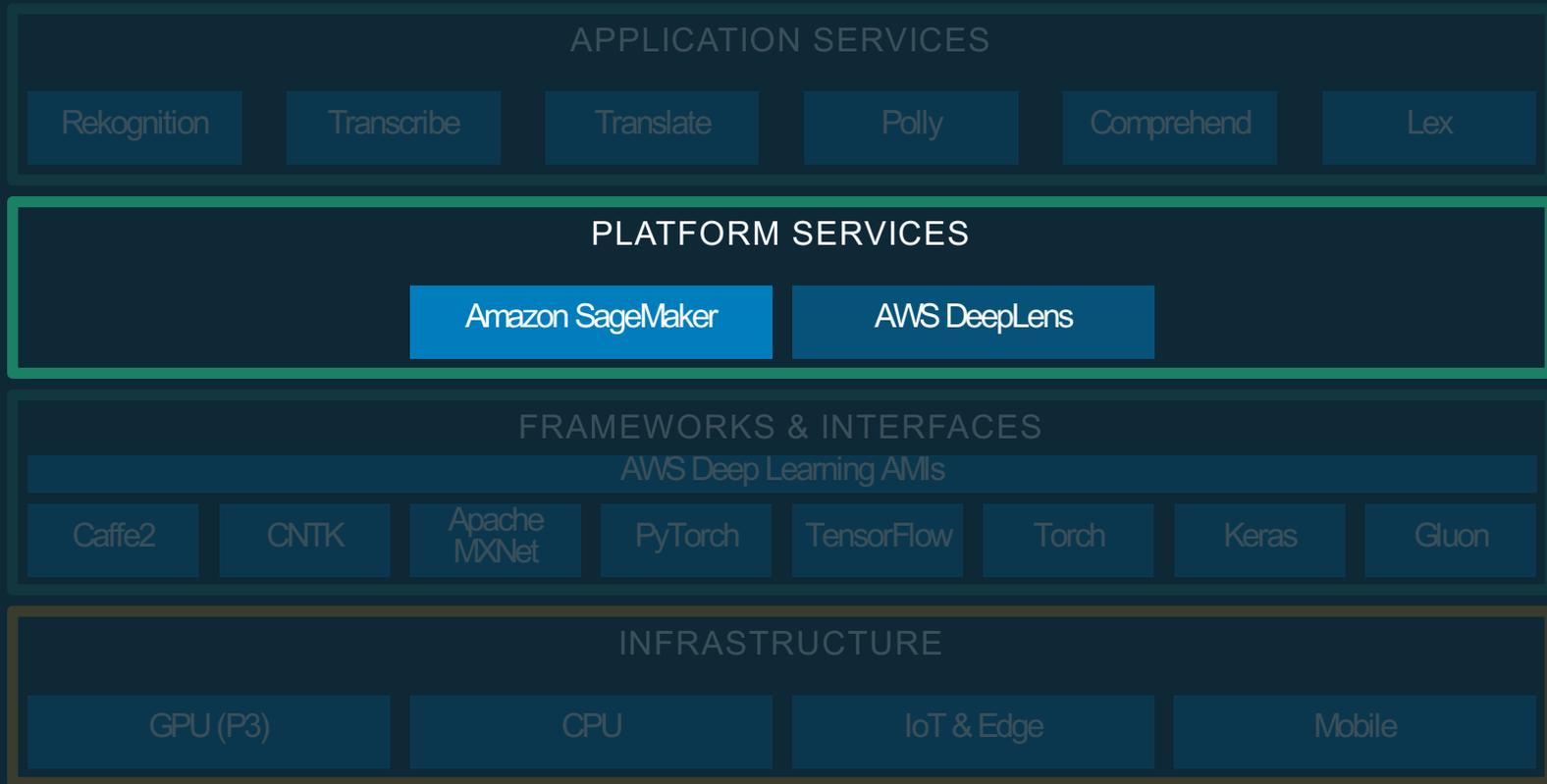
End-to-End Managed ML Platform

Lee Pang, Kevin Jorissen

The Amazon AI/DL/ML Stack



Machine Learning Platforms





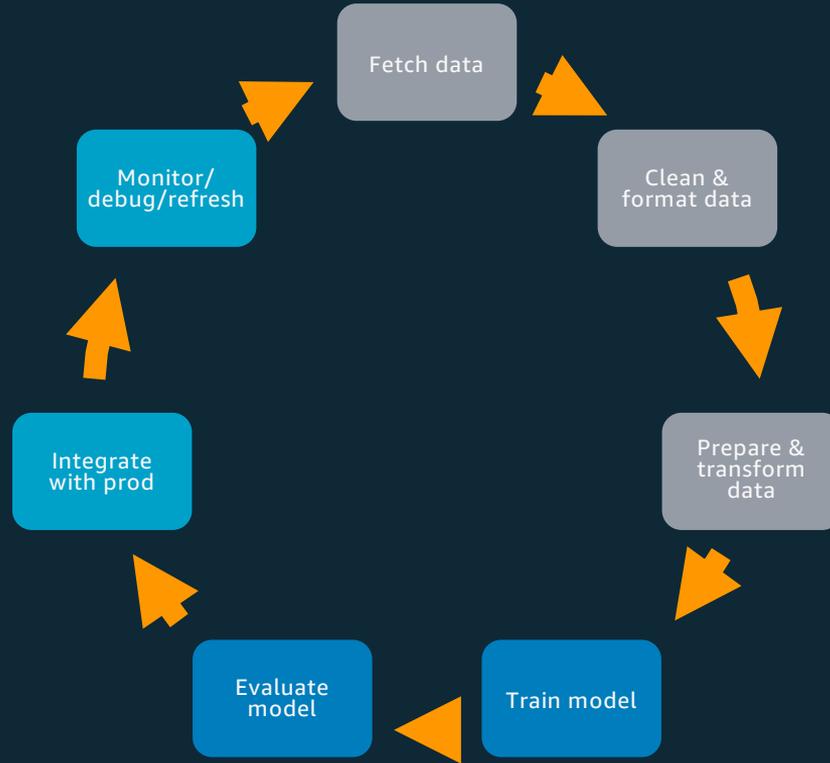
Amazon SageMaker

A **fully managed service** that enables **data scientists** and **developers** to quickly and easily **build** machine-learning based models **into production** smart applications.

Machine learning process is hard...

3. Deployment

- Setup and manage inference clusters
- Manage and auto scale inference APIs
- Testing, versioning, and monitoring



1. Data wrangling

- Setup and manage Notebook environments
- Get data to notebooks securely

2. Experimentation

- Setup and manage clusters
- Scale/distribute ML algorithms

Amazon SageMaker

Build, train, and deploy machine learning models at scale



End-to-End
Machine Learning
Platform



Zero setup



Flexible Model
Training



Pay by the second



Amazon SageMaker

1



Notebook Instances

2



Algorithms

3



ML Training Service

4



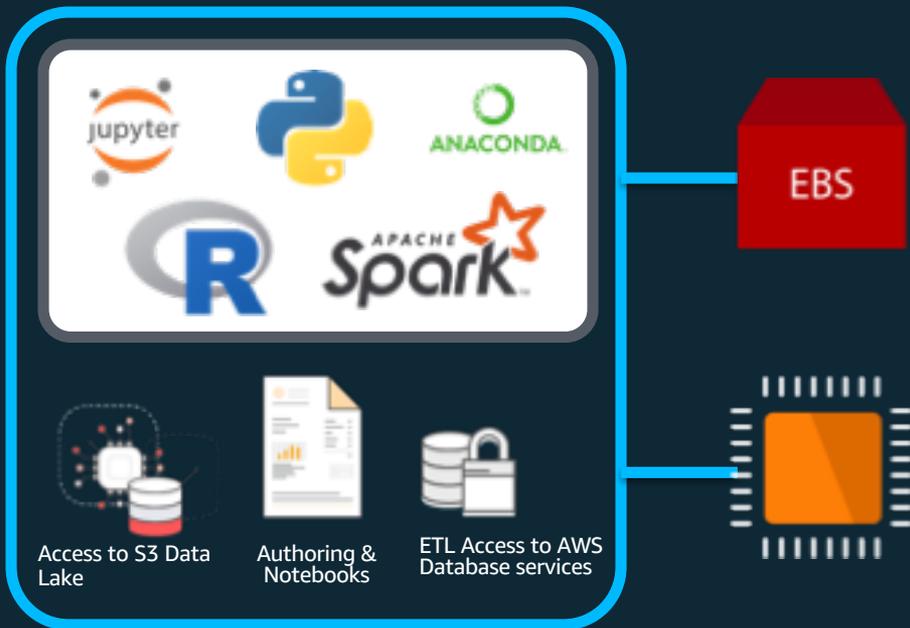
ML Hosting Service

1

Zero Setup For Exploratory Data Analysis



Notebook Instances



"Just add data"

- Recommendations/Personalization
- Fraud Detection
- Forecasting
- Image Classification
- Churn Prediction
- Marketing Email/Campaign Targeting
- Log processing and anomaly detection
- Speech to Text
- More...

2

Amazon SageMaker: 10x better algorithms



Algorithms



- Matrix Factorization
- Regression
- Principal Component Analysis
- K-Means Clustering
- Gradient Boosted Trees
- And More!

Amazon provided Algorithms



Bring Your Own Script (SM builds the Container)



SM Estimators in Apache Spark



Bring Your Own Algorithm (You build the Container)



Streaming datasets, for cheaper training



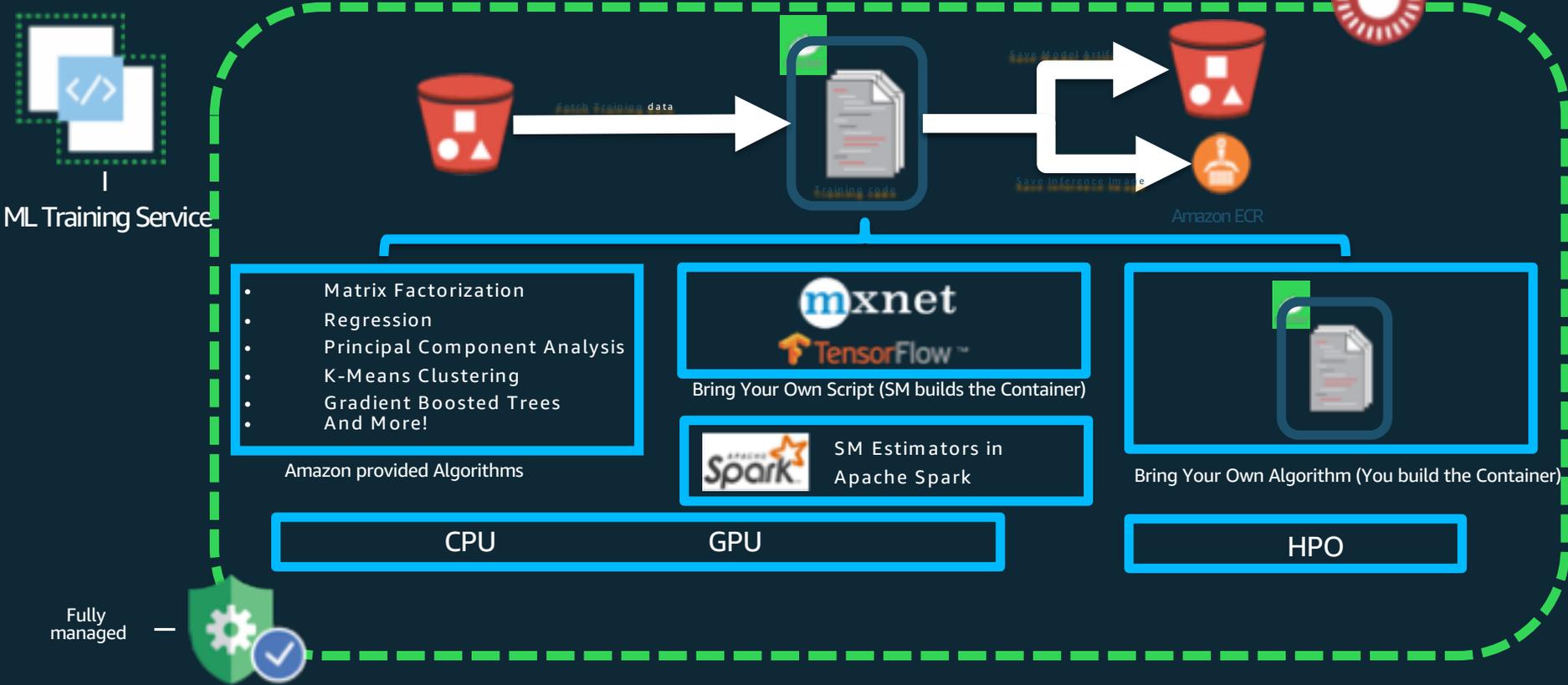
Train faster, in a single pass



Greater reliability on extremely large datasets

3

Managed Distributed Training with Flexibility

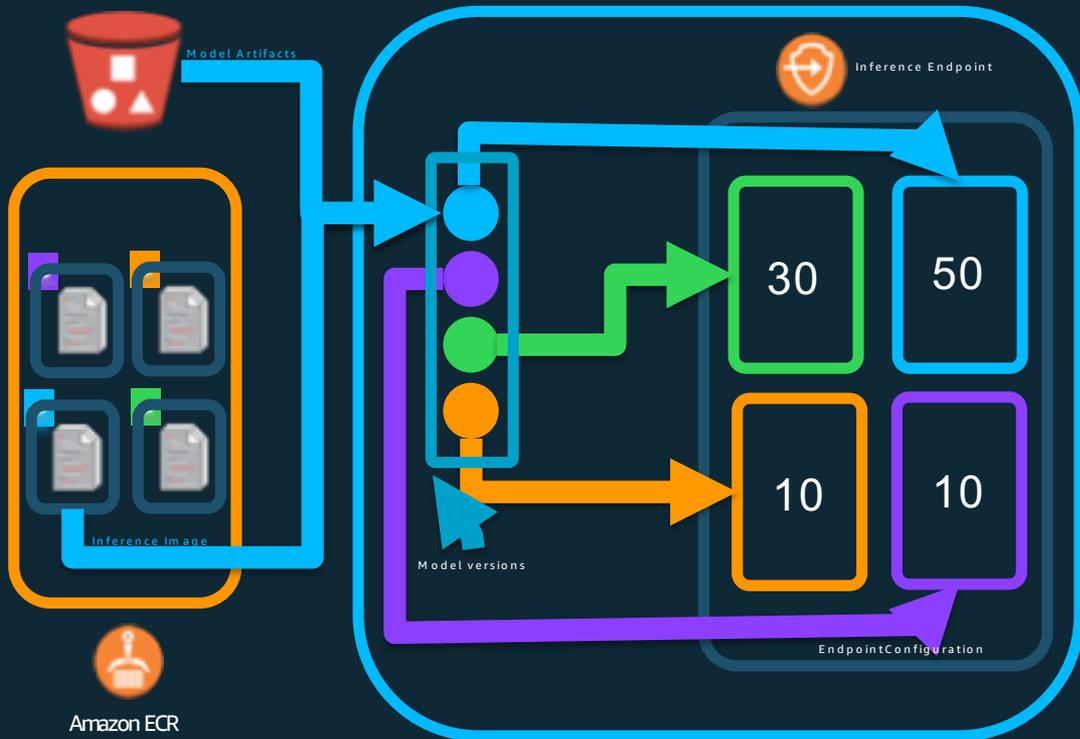


4

Easy Model Deployment to Amazon SageMaker



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



```
InstanceType: c3.4xlarge
InitialInstanceCount: 3
ModelName: prod
VariantName: primary
InitialVariantWeight: 50
```

ProductionVariant

One-Click!



4

Easy Model Deployment to Amazon SageMaker



ML Hosting Service

- ✓ Auto-Scaling Inference APIs
- ✓ A/B Testing (more to come)
- ✓ Low Latency & High Throughput
- ✓ Bring Your Own Model
- ✓ Python SDK



Amazon SageMaker



Let's get started!

Learning Objectives

- End-to-End machine learning with SageMaker
- Deep learning frameworks and distributed training
- Bringing your own model
- Leveraging public datasets

Prerequisites

AWS Account

Your own (recommended) with a user or role with full permissions to:

- AWS IAM
- Amazon S3
- Amazon SageMaker

AWS Region

Choose one of the following for all resources created in this workshop:

- **Oregon (us-west-2)**
- **N. Virginia (us-east-1)**
- Ohio (us-east-2)
- Ireland (eu-west-1)

Lab Content

Download from:

<https://bit.ly/2HhD2SG>

Setup

1. Create an S3 Bucket:
 1. Name: smworkshop-firstname-lastname
 2. Region: your region of choice
2. Launch a Notebook instance
 1. Region: your region of choice
 2. Instance Type: ml.m4.xlarge
 3. IAM role: "Create a new role"
 4. S3 Bucket: (the one you created above)



Lab 1

Introduction to Amazon SageMaker and Amazon Algorithms



Amazon SageMaker – End to End



Pre-built
notebooks for
common
problems



Built-in, high
performance
algorithms



One-click
training



Hyperparameter
optimization



One-click
deployment



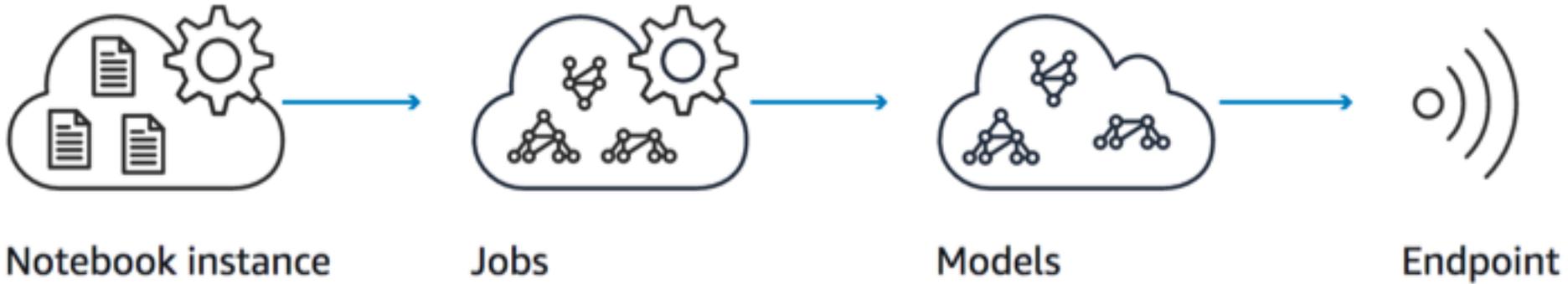
Fully managed
hosting with auto-
scaling

BUILD

TRAIN

DEPLOY

Amazon SageMaker – End to End



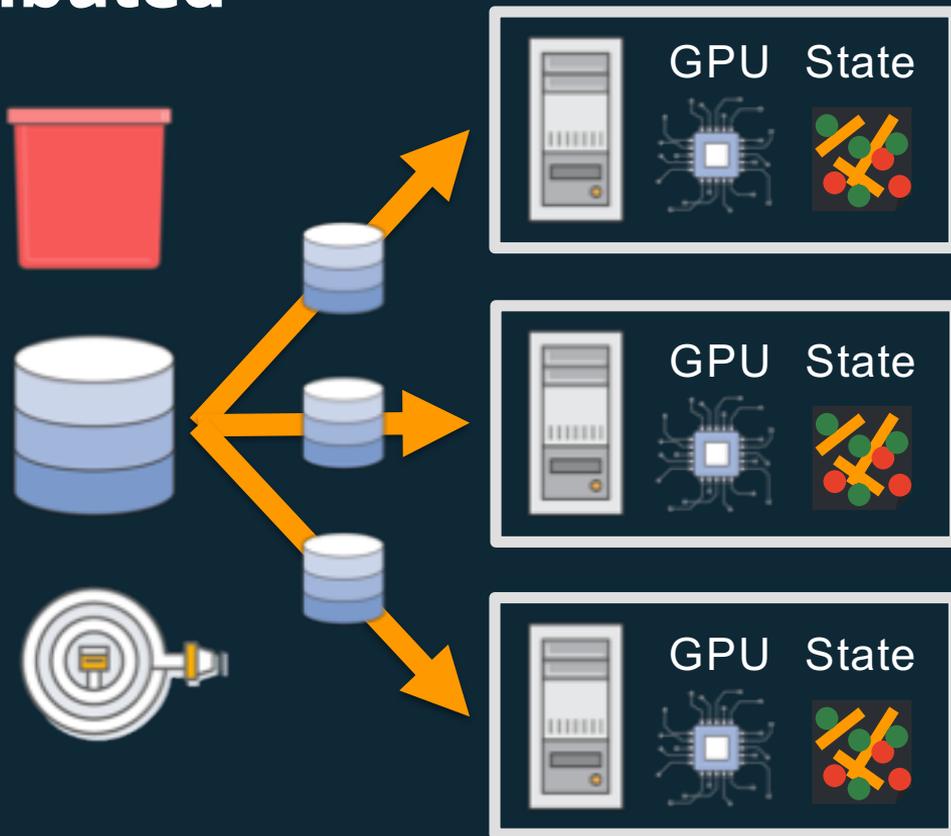


Lab 2

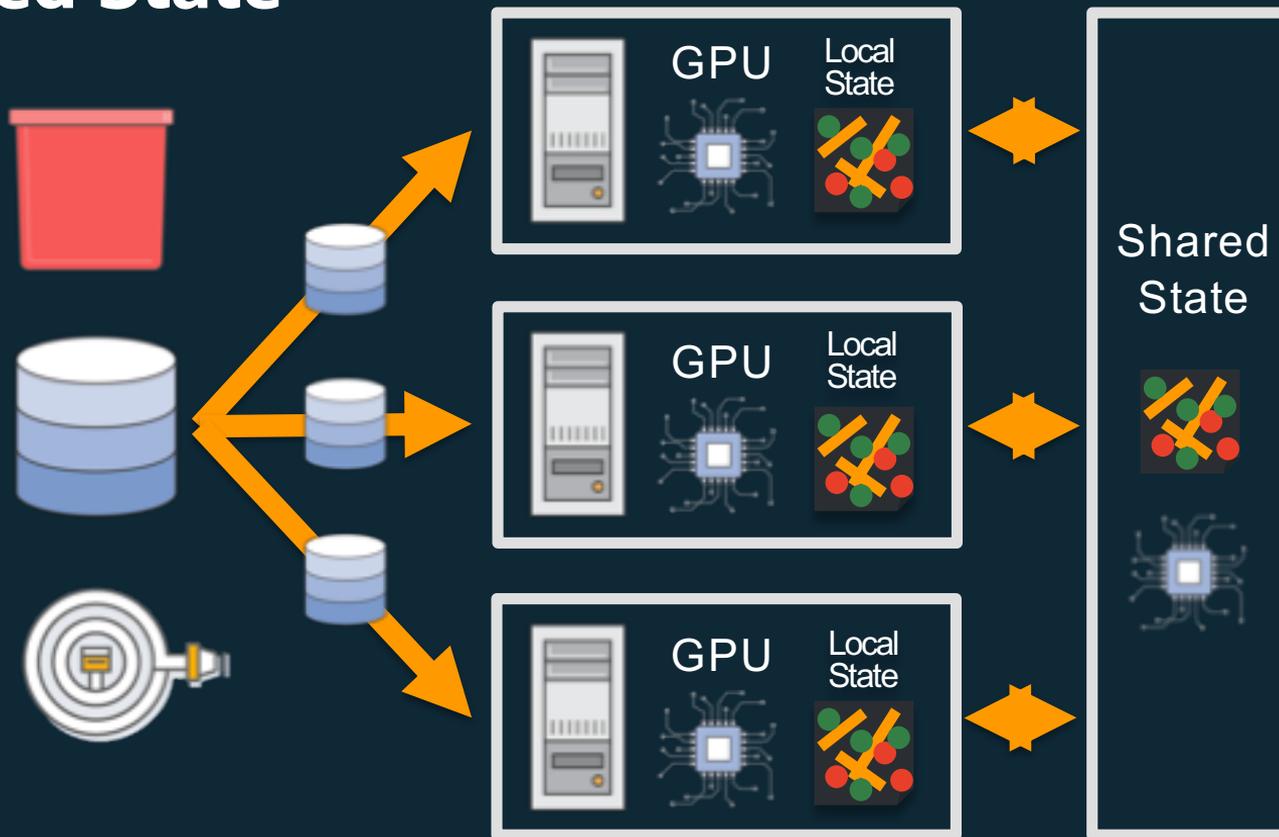
Distributed Training with TensorFlow



Distributed



Shared State





Lab 3

Bringing Your Own Algorithms



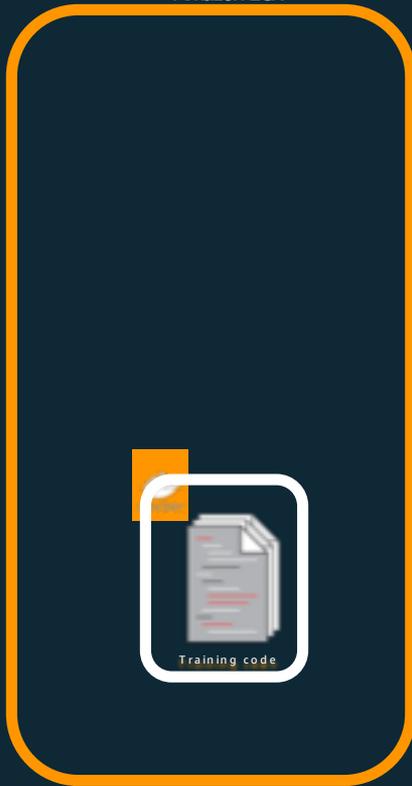
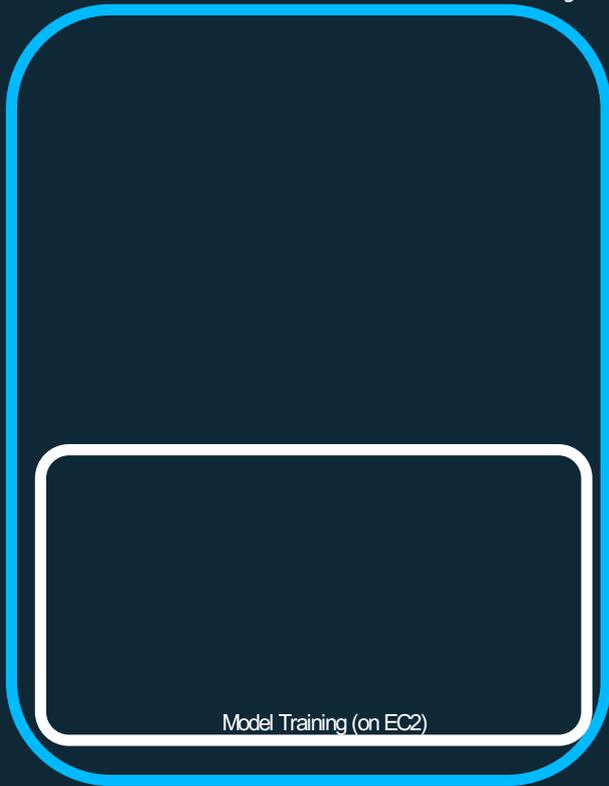
Client application



Amazon SageMaker



Amazon ECR



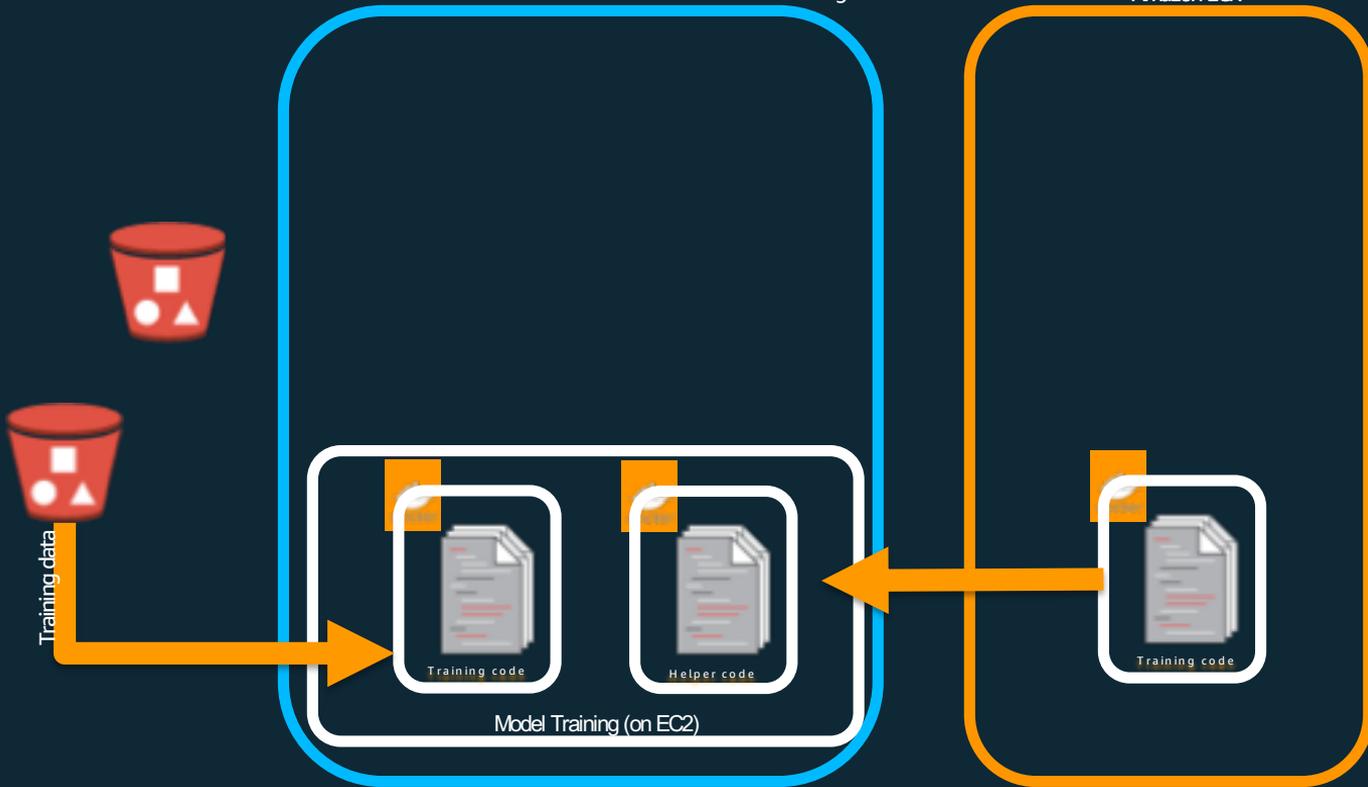
Client application



Amazon SageMaker



Amazon ECR



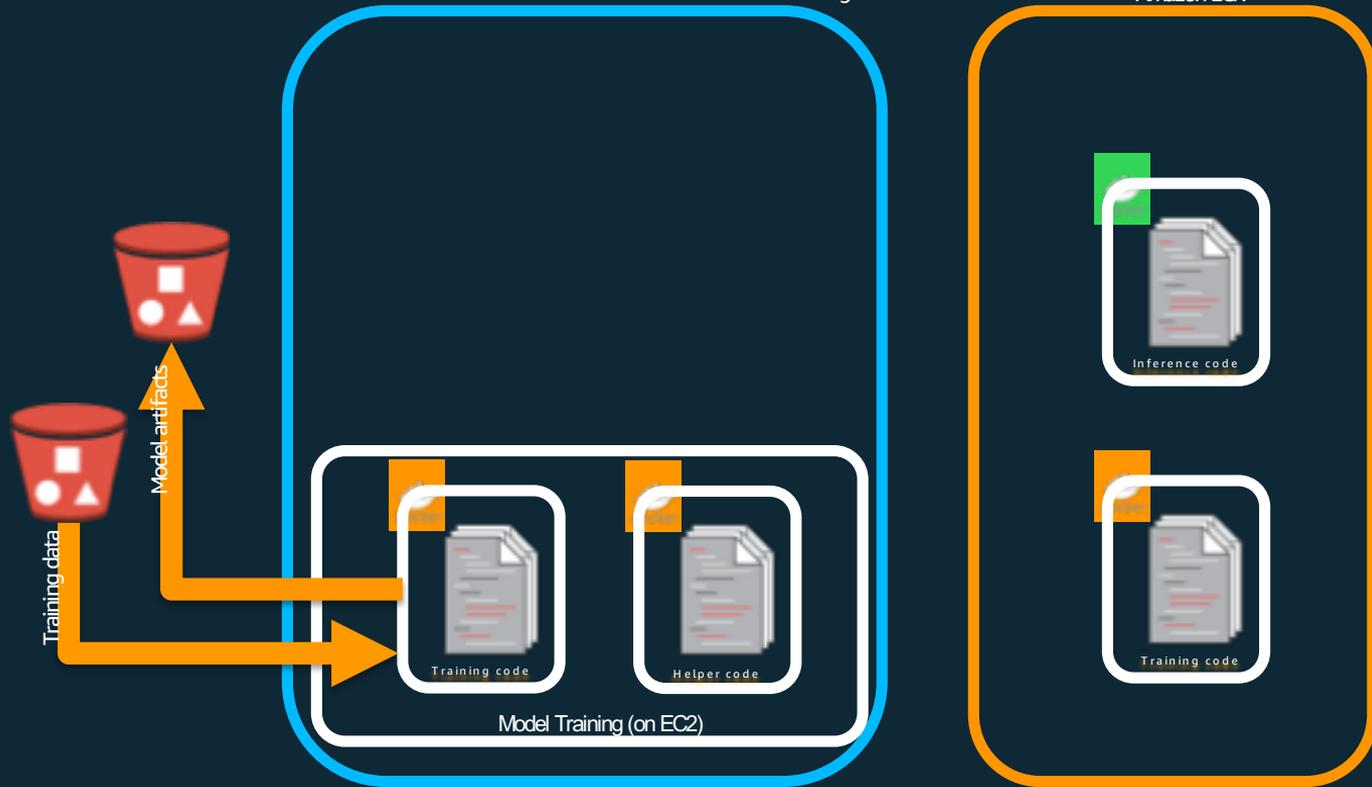
Client application



Amazon SageMaker



Amazon ECR



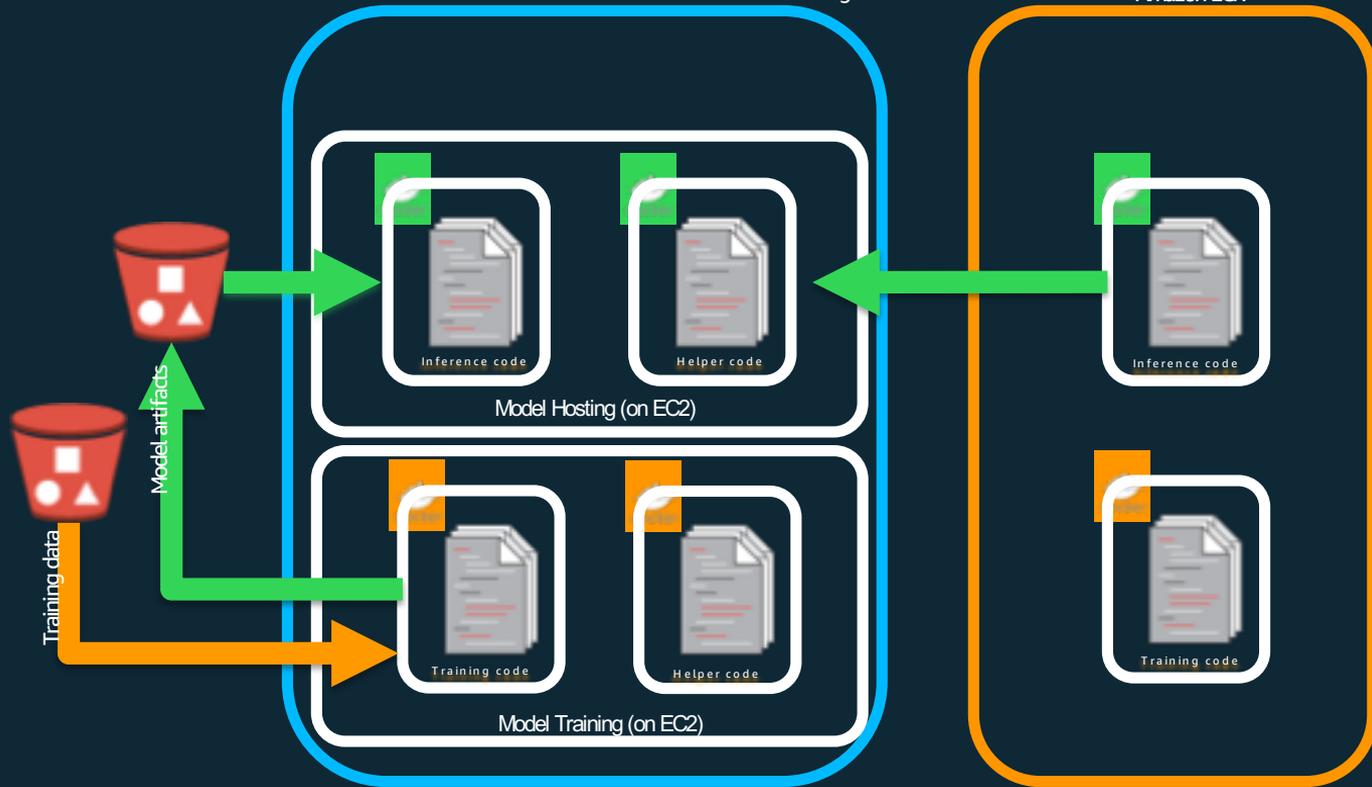
Client application

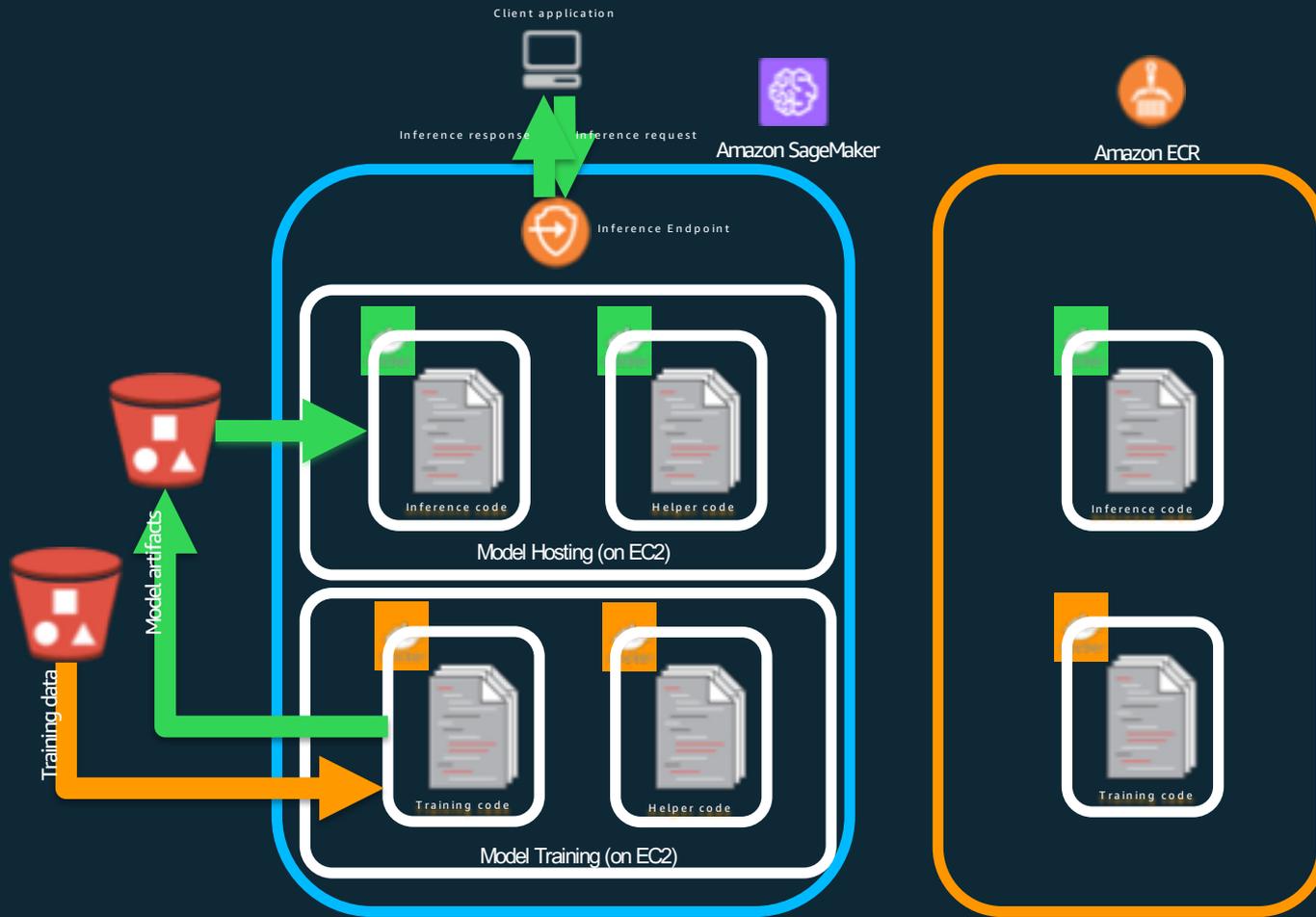


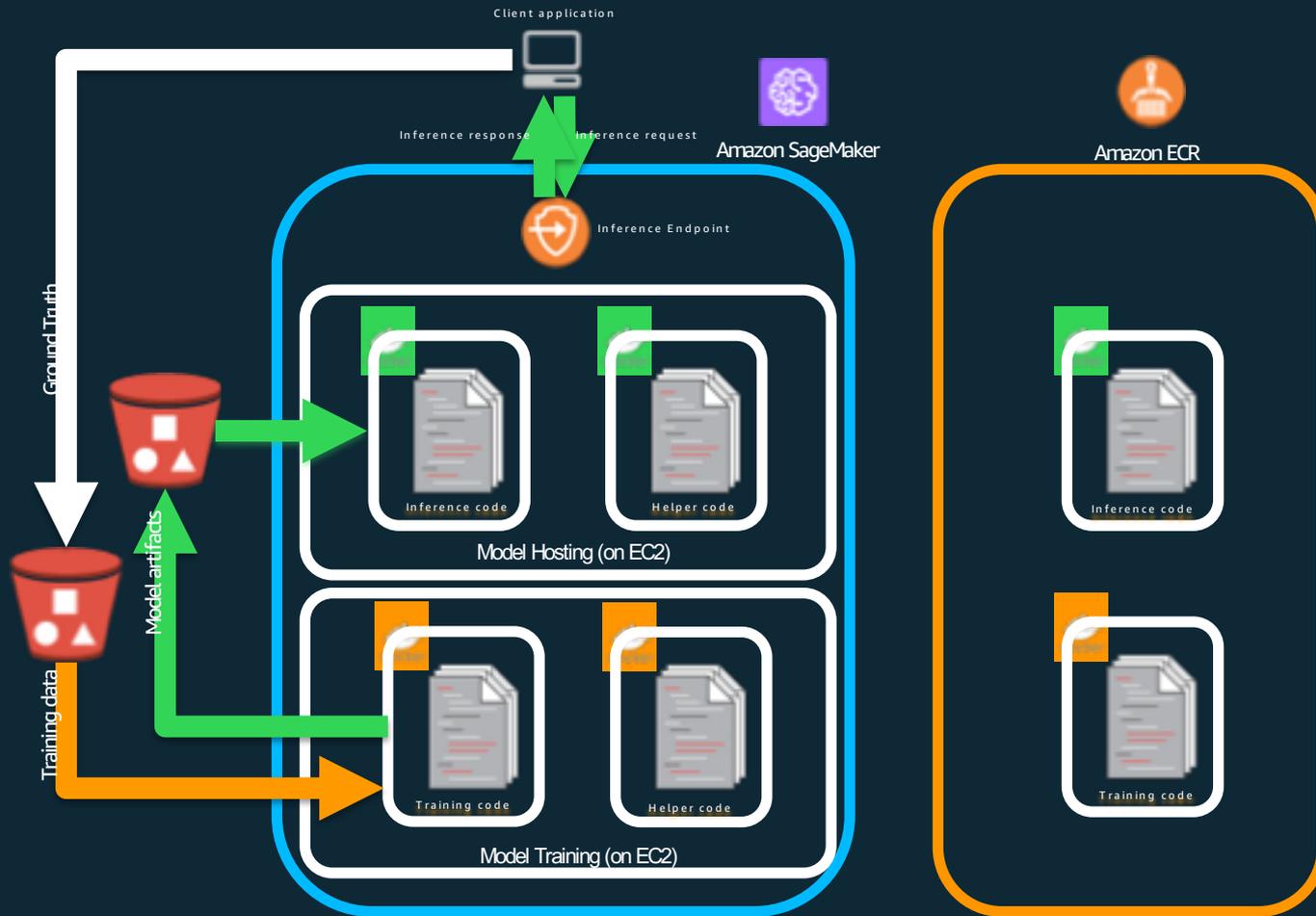
Amazon SageMaker



Amazon ECR









Lab 4

Using Public Datasets



Public Data on AWS

AWS hosts a variety of public datasets that anyone can access for free.

<https://aws.amazon.com/public-datasets/>

1000 Genomes Project

The 1000 Genomes Project is an international collaboration which has established the most detailed catalogue of human genetic variation, including SNPs, structural variants, and their haplotype context.

<https://aws.amazon.com/1000genomes/>

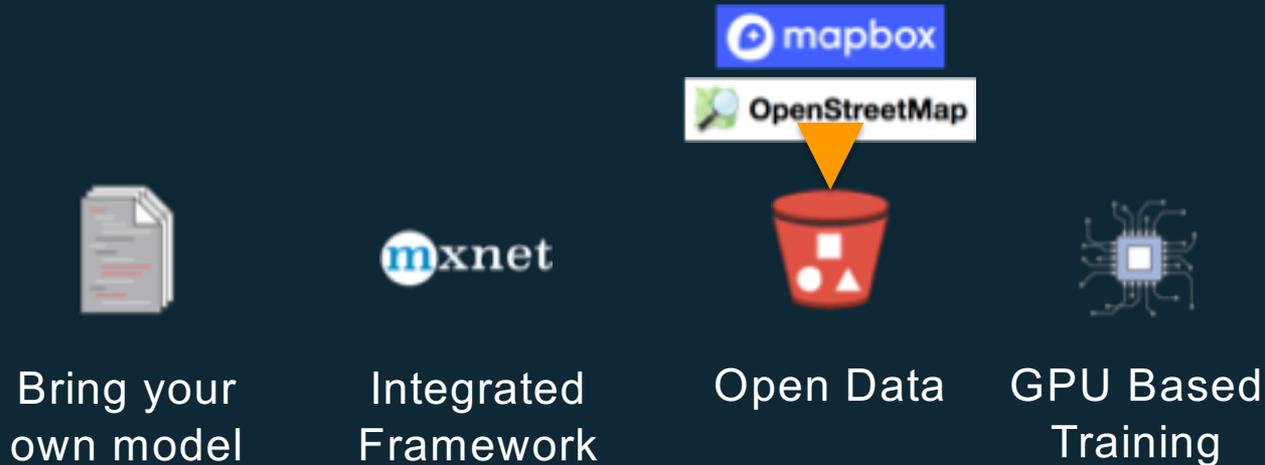


Lab 5

Classifying Buildings in Vietnam
MXNet, GPU instances, and Open Map Data



A real world example



Developed by [developmentSEED.org](https://developmentseed.org)

<https://developmentseed.org/blog/2018/01/19/sagemaker-label-maker-case/>

Clean-up!

Avoid charges for resources you no longer need after this workshop

- Endpoints
- Notebook instances
- S3 Bucket

Review

- ✓ End-to-End machine learning with SageMaker
 - Linear Learner binary classification of MNIST
- ✓ Deep learning frameworks and distributed training
 - TensorFlow CNN on MNIST
- ✓ Bringing your own model
 - Deploying scikit-learn decision trees
- ✓ Leveraging public datasets
 - K-means clustering of 1000 Genomes data

Amazon SageMaker Resources

- Getting started with Amazon SageMaker: <https://aws.amazon.com/sagemaker/>
- Use the Amazon SageMaker SDK:
 - For Python: <https://github.com/aws/sagemaker-python-sdk>
 - For Spark: <https://github.com/aws/sagemaker-spark>
- SageMaker Examples: <https://github.com/aws-labs/amazon-sagemaker-examples>
- Let us know what you build!