AWS | Introduction to Amazon SageMaker

The Amazon ML stack

AI SERVICES

Vision


REKOGNITION IMAGE


REKOGNITION VIDEO


TEXTRACT

Speech


POLLY


TRANSCRIBE

Language


TRANSLATE


COMPREHEND

Chatbots


LEX

Forecasting


FORECAST

Recommendations


PERSONALIZE

ML SERVICES


AMAZON SAGEMAKER

BUILD

- Pre-built algorithms & notebooks
- Data labeling (GROUND TRUTH)
- Algorithms & models (AWS MARKETPLACE FOR MACHINE LEARNING)

TRAIN

- One-click model training & tuning
- Optimization (NEO)
- Reinforcement learning

DEPLOY

- One-click deployment & hosting
- DeepRacer
- DeepLens

ML FRAMEWORKS & INFRASTRUCTURE

Frameworks & Interfaces

Caffe2 · CNTK · Keras · mxnet · PYTORCH · GLUON · TensorFlow · torch · Chainer

Infrastructure


EC2 P3 & P3N


EC2 C5

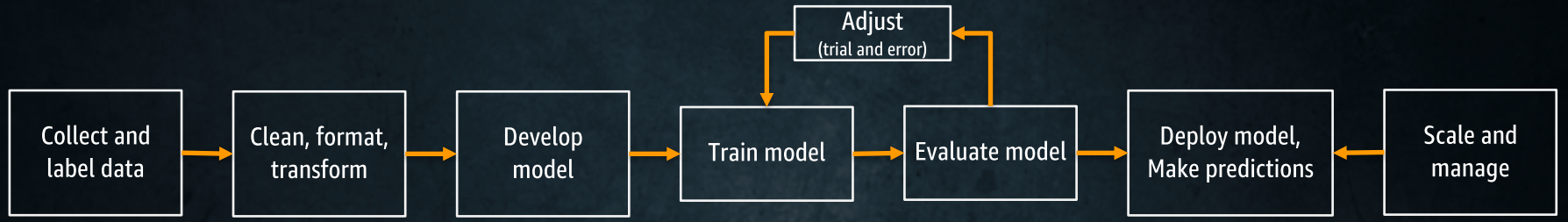

FPGAs


GREENGRASS


ELASTIC INFERENCE

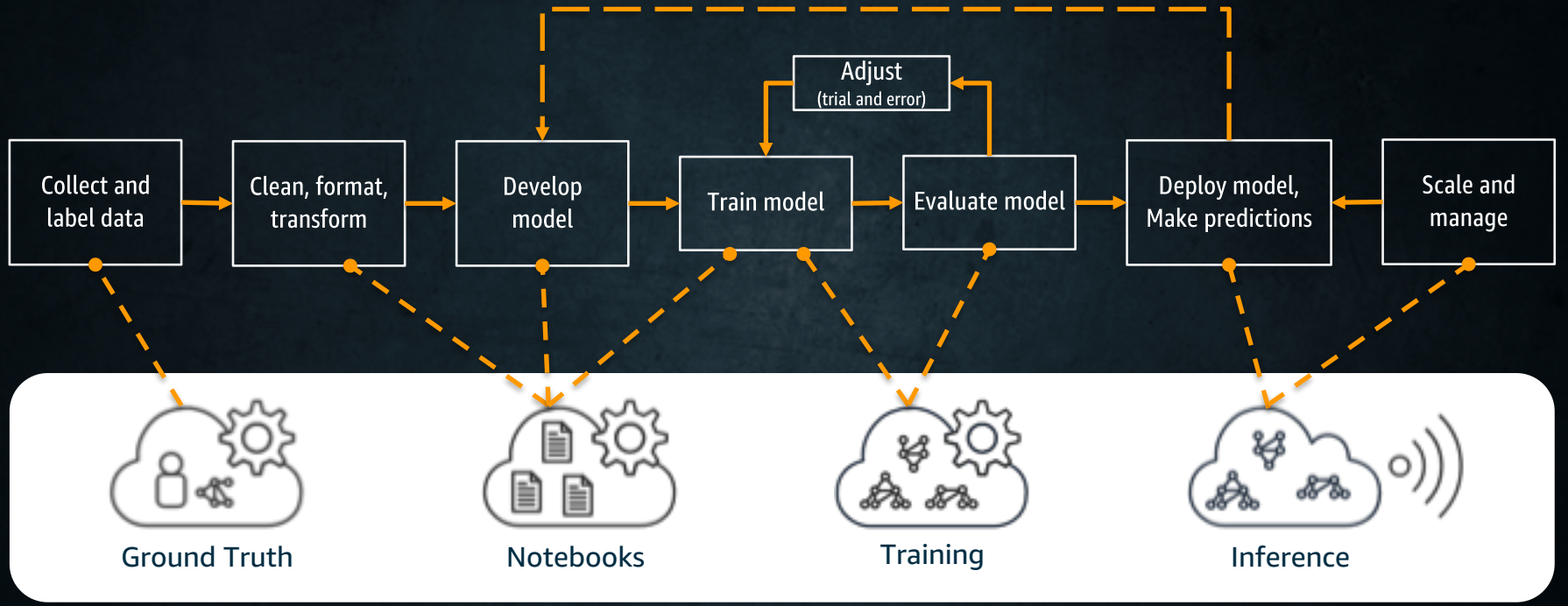
aws

ML was too complicated for everyday developers



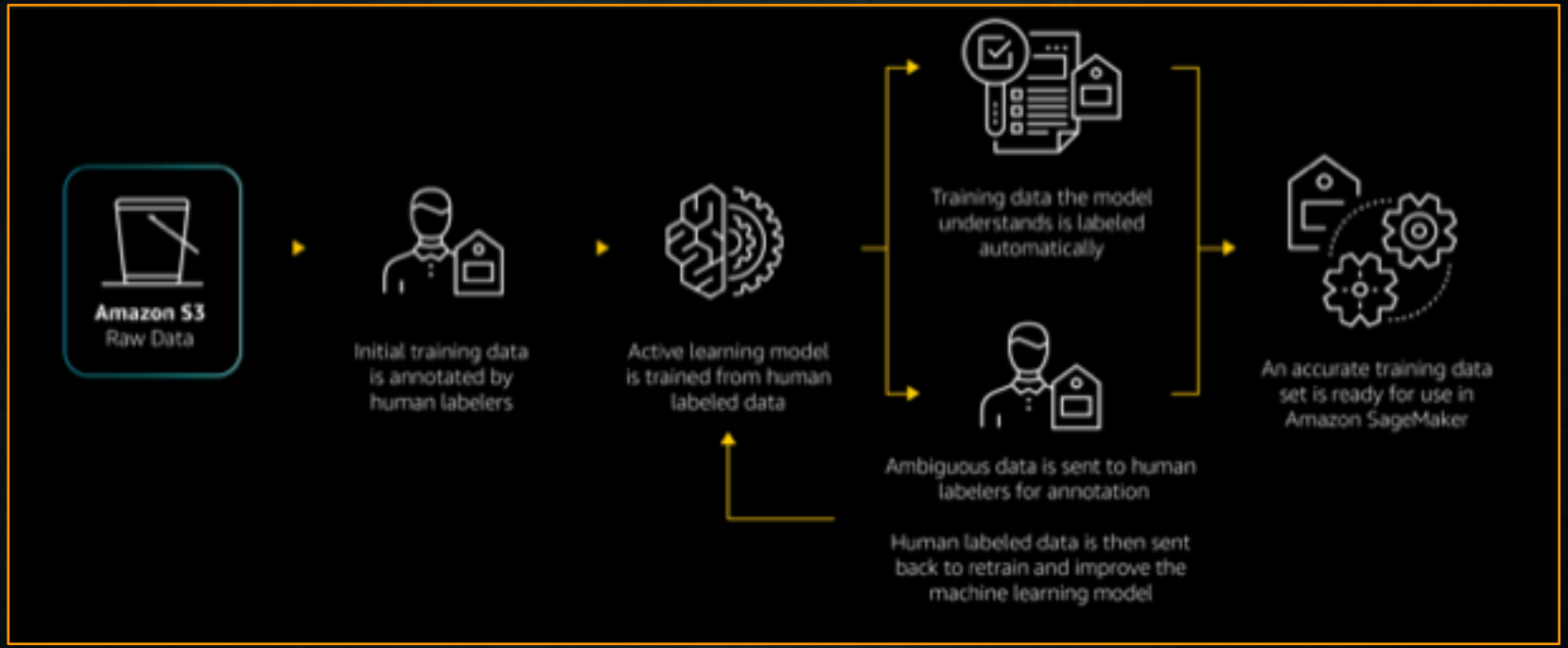
Amazon SageMaker

Build, train, and deploy machine learning models quickly

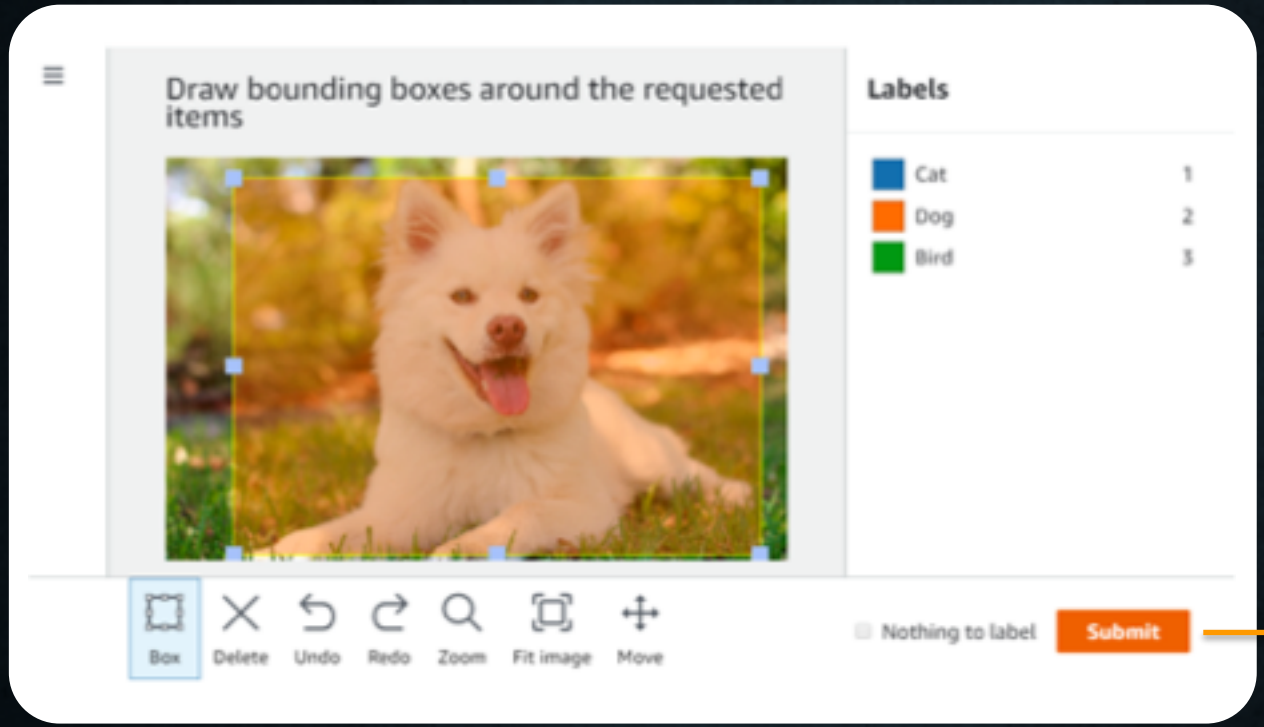


Amazon SageMaker | Ground Truth

Amazon SageMaker | Ground Truth



Amazon SageMaker | Ground Truth




```
[
  {
    "boundingBox": {
      "boundingBoxes": [
        {
          "height": 291,
          "label": "dog",
          "left": 59,
          "top": 86,
          "width": 417
        }
      ],
      "inputImageProperties": {
        "height": 480,
        "width": 640
      }
    }
  }
]
```

Amazon SageMaker | Ground Truth

Creating training data

Mechanical Turk workers

Private labeling workforce

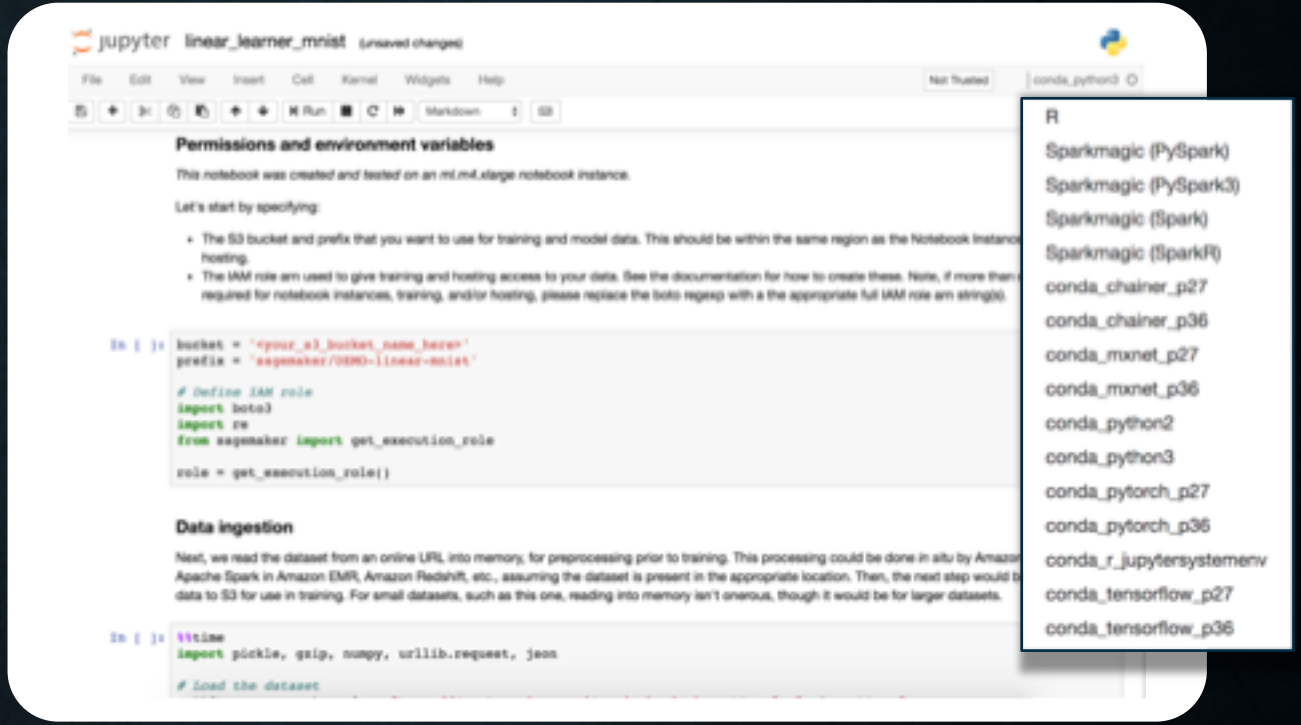
Third-party vendors

aws

Amazon SageMaker | Notebooks

Amazon SageMaker | Notebooks

- Jupyter notebooks
- Support Jupyter Lab
- Multiple built-in kernels
- Install external libraries and kernels
- Integrate with Git
- Sample notebooks

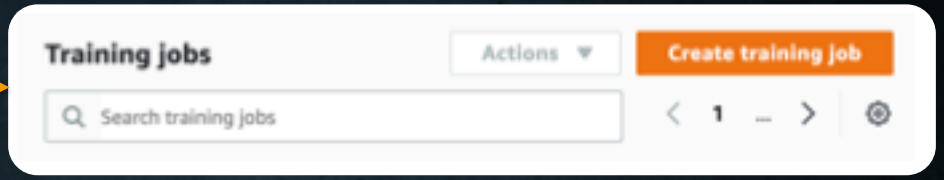


Amazon SageMaker | Training

Amazon SageMaker | Training

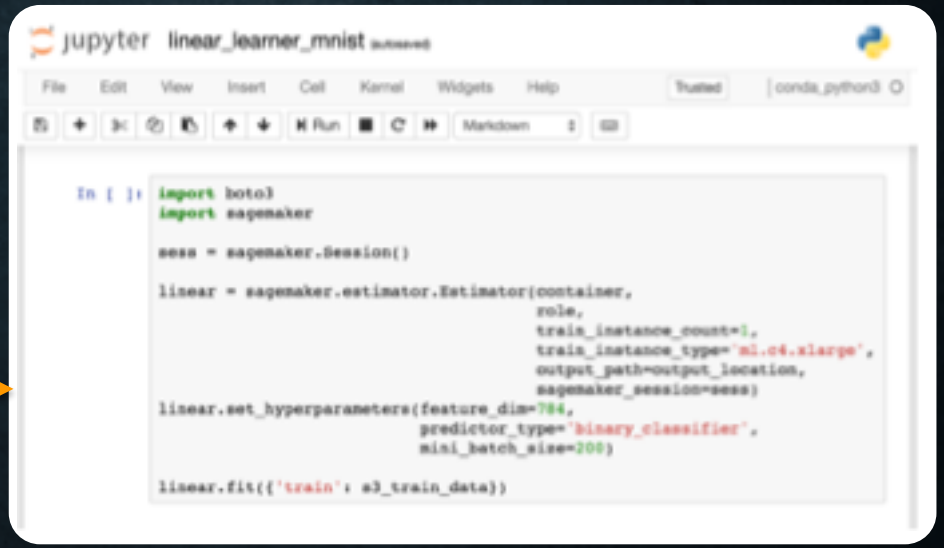
Launch Training

One click in Console



- OR -

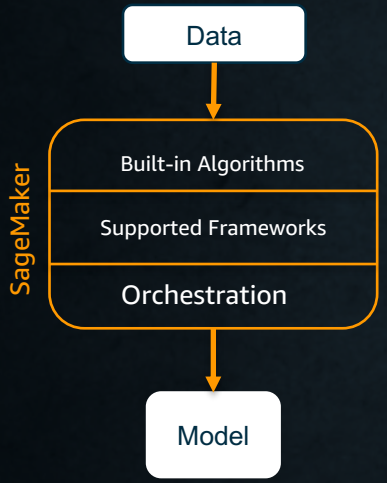
Using API/SDK



aws

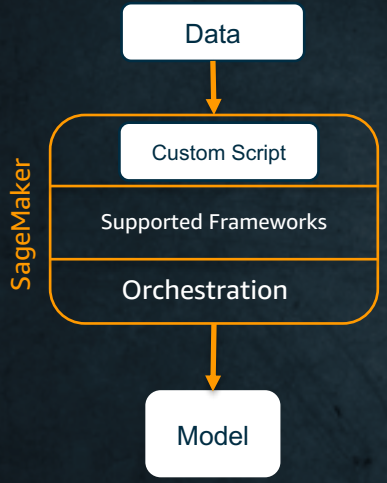
Amazon SageMaker | Training

Sagemaker Built-in algorithms



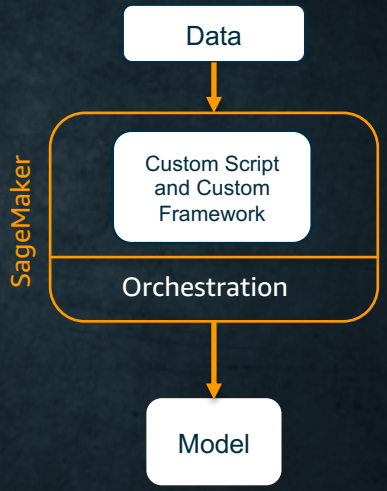
17 Built in high performance algorithms

Custom script on supported framework



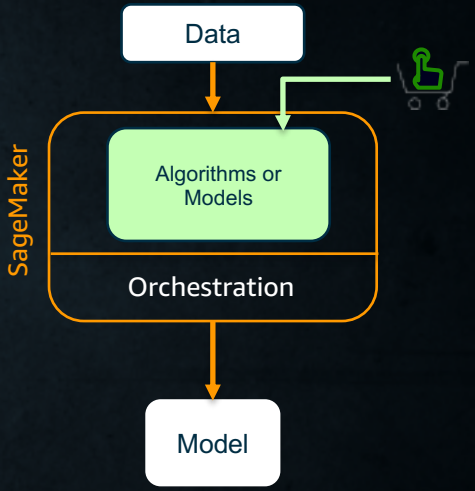
Supported Frameworks: Apache MXNet, TensorFlow , Scikit-learn, PyTorch, Chainer

BYO algorithm and framework



Docker containers with your own algorithms and frameworks

AWS Marketplace algorithms



3rd party algorithms and models

Amazon SageMaker | Training

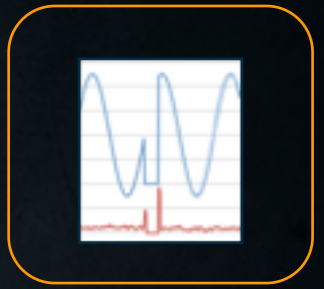
Built-in algorithms

XGBoost, FM, Linear, and Forecasting for supervised learning

Kmeans, PCA, and BlazingText (Word2Vec) for clustering and pre-processing

Image classification and object detection with convolutional neural networks

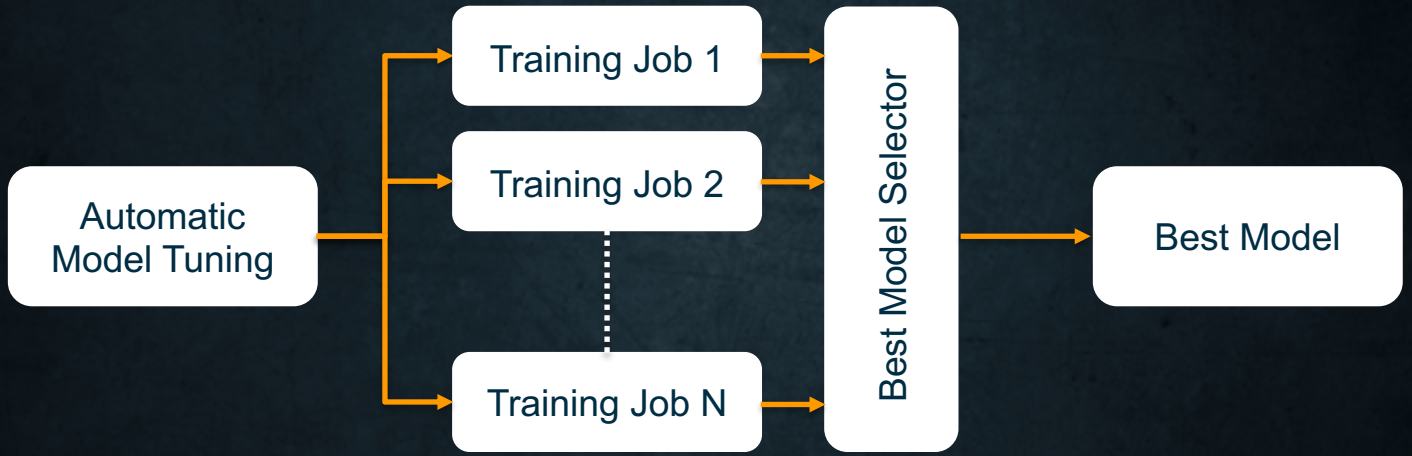
LDA and NTM for topic modeling, seq2seq for translation



Random Cut Forest for anomaly detection

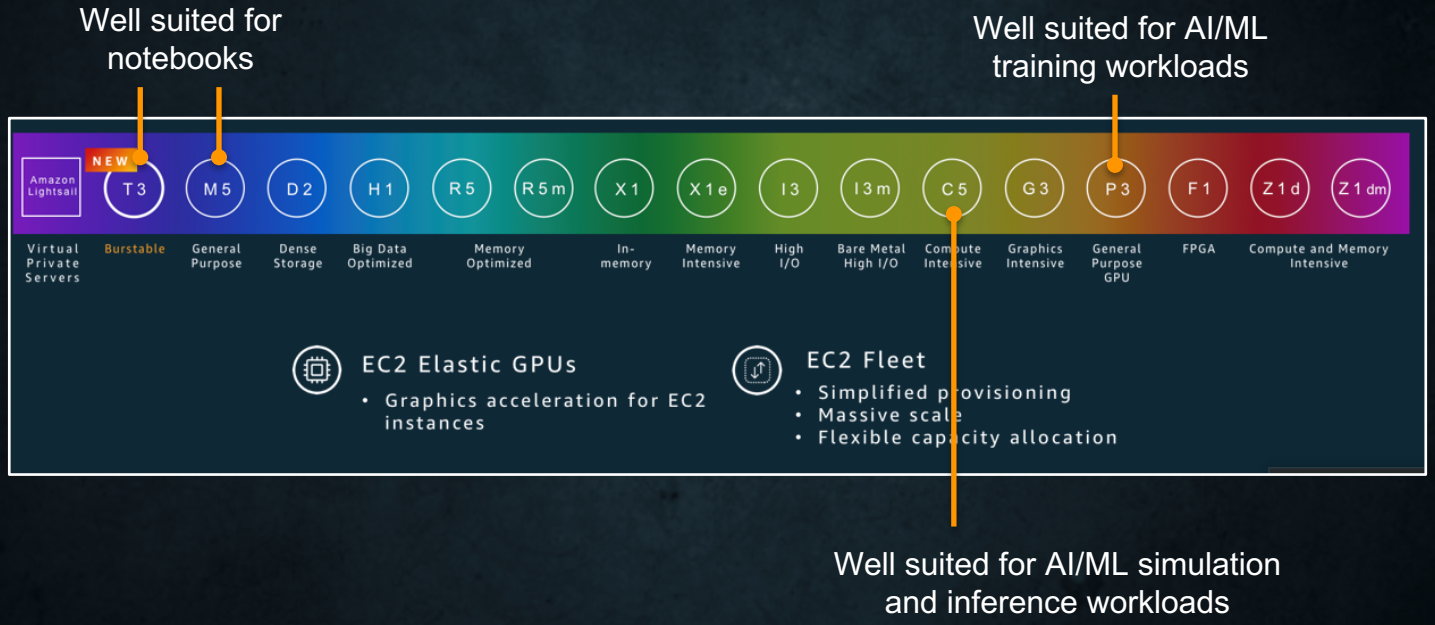
Amazon SageMaker | Training

Hyperparameter Optimization



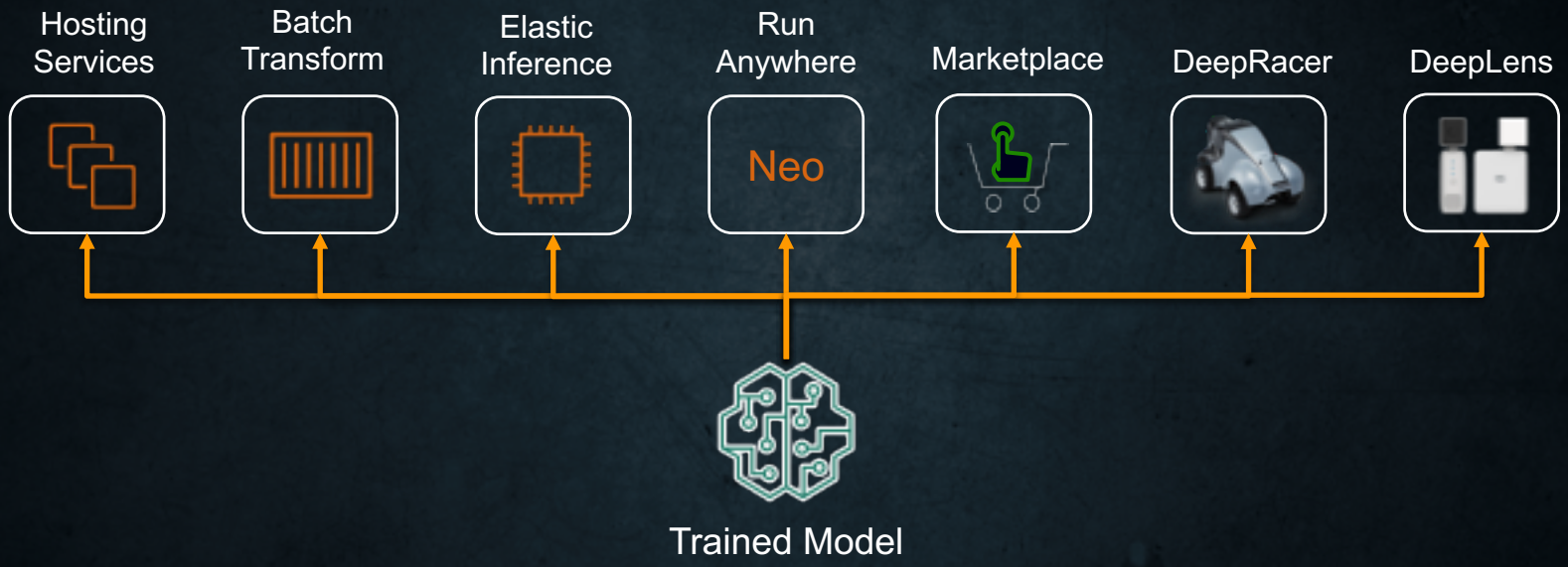
- Define Metrics
- Hyperparameter ranges/scaling
- Stop tuning job early
- Use warm start

Amazon SageMaker | Training



Amazon SageMaker | Inference

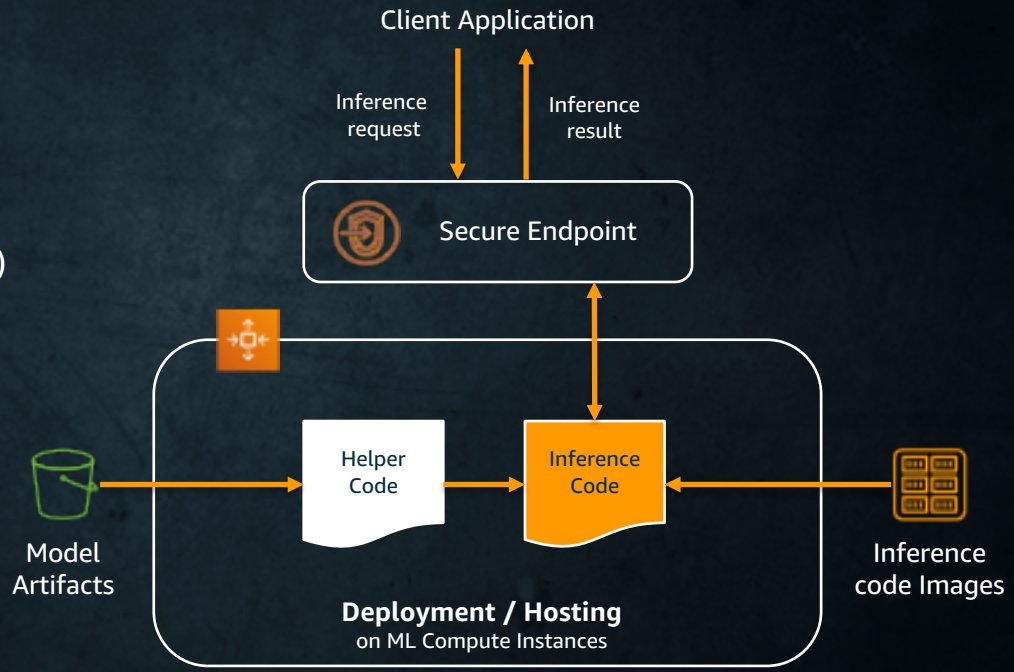
Amazon SageMaker | Inference



Amazon SageMaker | Inference

Hosting Services

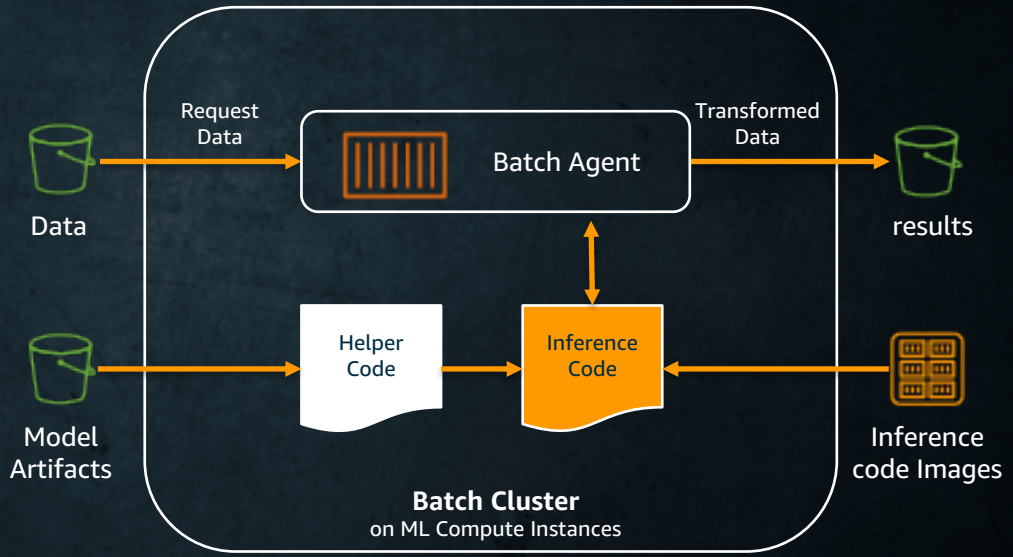
- Secure HTTPS endpoint
- Elastic configuration (scales with traffic)
- Multiple versions of the a model (A/B testing)
- Multi-AZ support
- Sub-second latencies
- Persistent deployments



Amazon SageMaker | Inference

Batch Transform

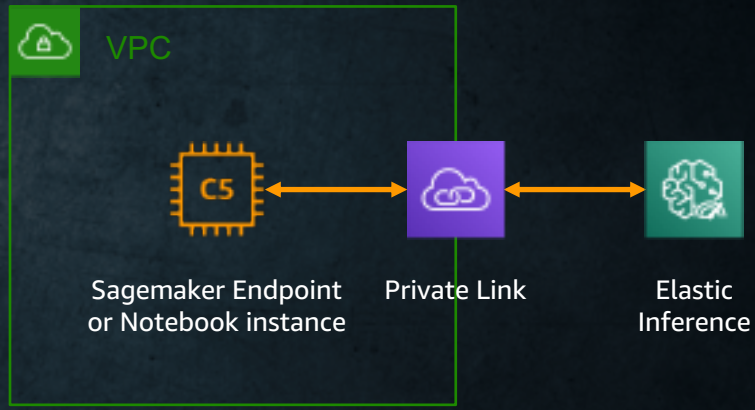
- Predictions for an entire dataset
- Transient resources (instances provisioned and terminated once job is done)
- No infrastructure to manage



Amazon SageMaker | Inference

Elastic Inference

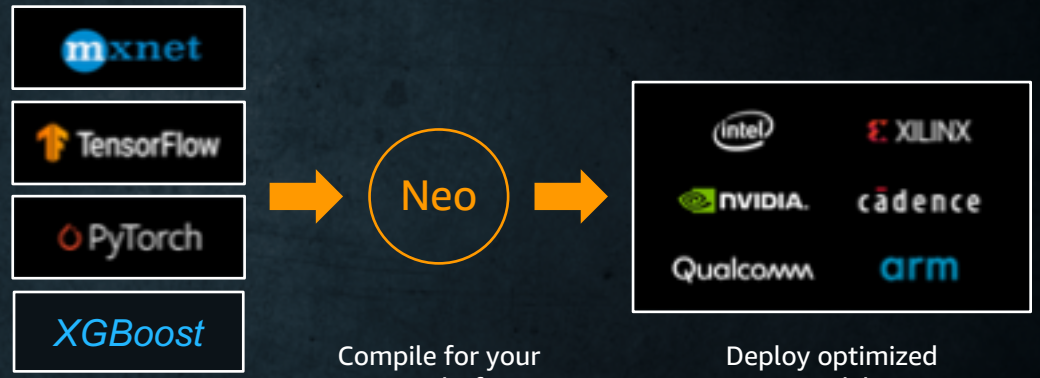
- Inference acceleration at a fraction of full GPU instance cost
- Add accelerators to CPU instances
- Works with inference as well as notebook instances
- Supports TensorFlow and MXNet frameworks, any other can be used via ONNX



Amazon SageMaker | Inference

Neo

- Train your model once and deploy anywhere
- Optimization for target platform
- Deploy on EC2, Edge, IoT
- Open Source
- Up to 2X performance increase
- 1/10th of the size of the original framework



Develop in your favorite framework

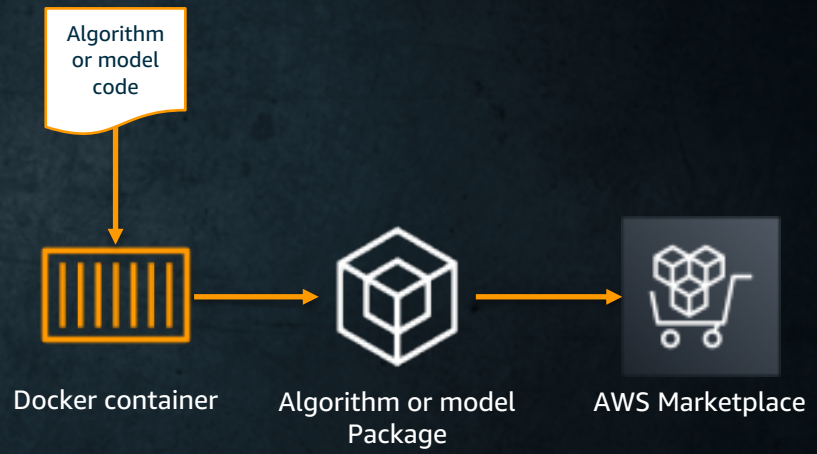
Compile for your target platform

Deploy optimized model

Amazon SageMaker | Inference

Marketplace

- Develop your algorithm or model, and package it in a Docker container.
- Create an algorithm or model package resource in Amazon SageMaker.
- Register as a seller on AWS Marketplace and list your algorithm or model package on AWS Marketplace.



Amazon SageMaker | Inference

DeepLens

Learn computer vision through projects, tutorials, and real-world hands-on exploration with the world's first deep learning enabled video camera for developers.

- Inference at the edge
- SageMaker trained models
- Simplified deployment
- Sample notebooks

DeepRacer

A fully autonomous, 1/18th scale race car, packed full of everything you need to learn about reinforcement learning through autonomous driving.

Amazon SageMaker | >10k users

aws

To learn more:

AWS Machine Learning University

Practical education on ML for new & experienced practitioners
Based on the same material used to train Amazon developers

ml.aws