DEV DAY

# Build, train and deploy Machine Learning models at scale

Julien Simon
Global Evangelist, AI & Machine Learning
@julsimon

aws

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# The Amazon ML Stack: Broadest & Deepest Set of Capabilities

## AI SERVICES

Vision | Speech | Language | Chatbots | Forecasting | Recommendations

REKOGNITION IMAGE

REKOGNITION VIDEO

TEXTRACT NEW

POLLY

TRANSCRIBE

TRANSLATE

COMPREHEND

COMPREHEND MEDICAL NEW

LEX

FORECAST NEW

PERSONALIZE NEW

## ML SERVICES

AMAZON SAGEMAKER

**BUILD**

Pre-built algorithms & notebooks NEW

Data labeling (GROUND TRUTH) NEW

Algorithms & models (AWS MARKETPLACE)

**TRAIN**

One-click model training & tuning

Optimization (NEO) NEW

Models without training data (REINFORCEMENT LEARNING)

**DEPLOY**

One-click deployment & hosting NEW

## ML FRAMEWORKS & INFRASTRUCTURE

Frameworks | Interfaces | Infrastructure

TensorFlow

mxnet

PYTORCH

GLUON

Keras

EC2 P3 & P3dn

EC2 C5

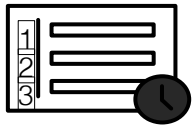
FPGAs

GREENGRASS

ELASTIC INFERENCE NEW

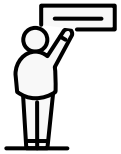
# Amazon SageMaker:
# Build, Train, and Deploy ML Models at Scale

Collect and prepare training data



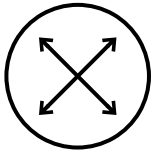

Choose and optimize your ML algorithm

Set up and manage environments for training



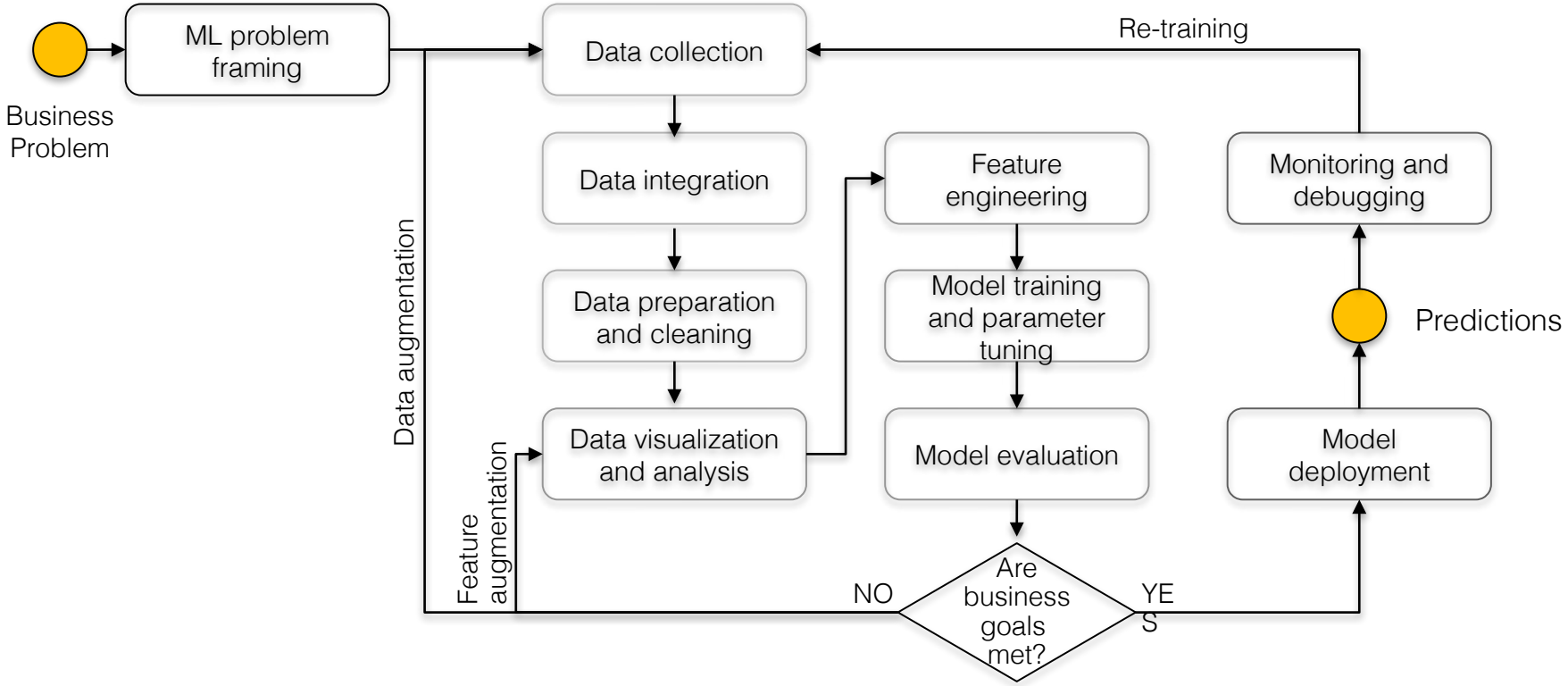
Train and Tune ML Models

Deploy models in production



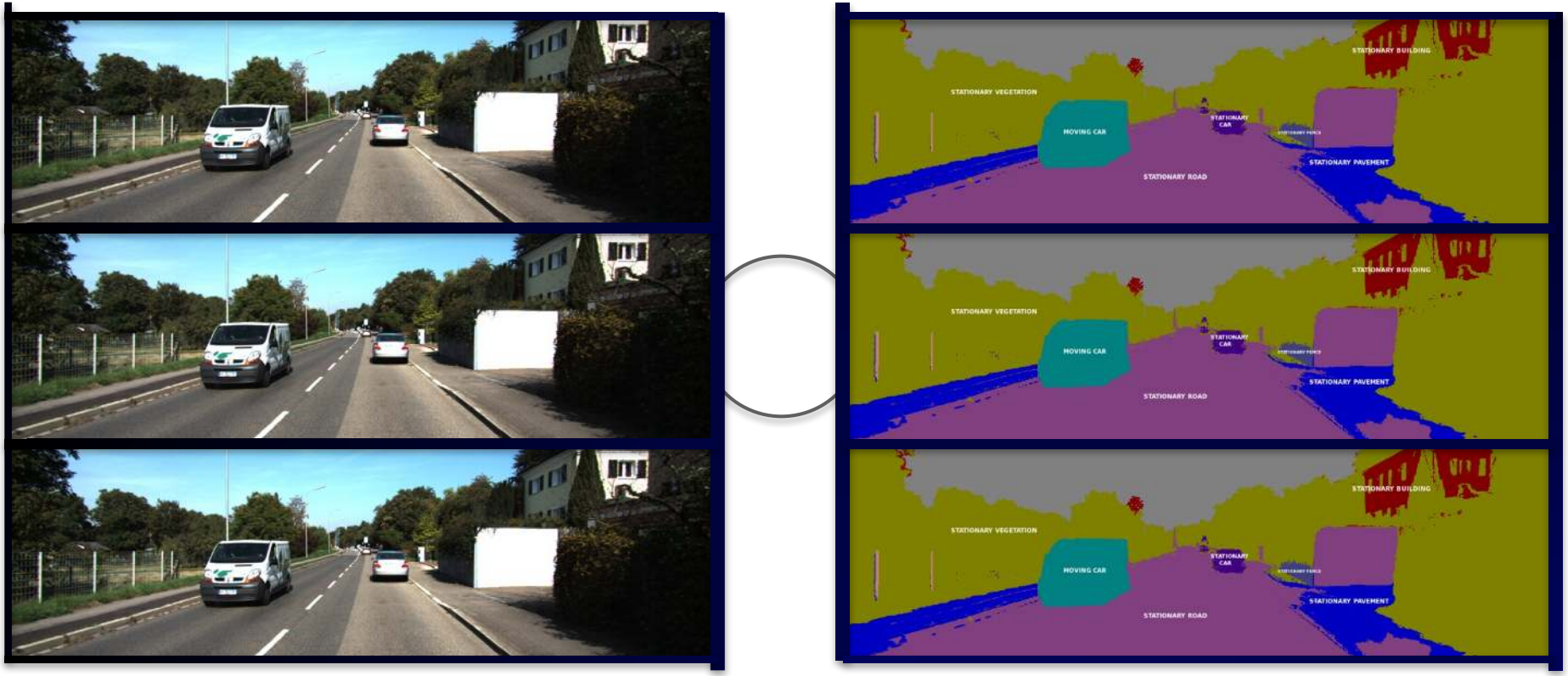
Scale and manage the production environment

# Machine learning cycle



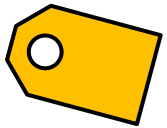
# Successful models require high-quality data

# Successful models require high-quality data

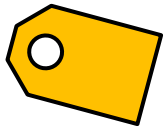


# Amazon SageMaker Ground Truth

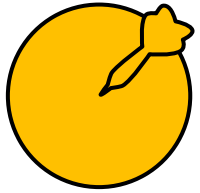
https://aws.amazon.com/blogs/aws/amazon-sagemaker-ground-truth-build-highly-accurate-datasets-and-reduce-labeling-costs-by-up-to-70



Quickly label training data



Easily integrate human labelers



Get accurate results

## KEY FEATURES

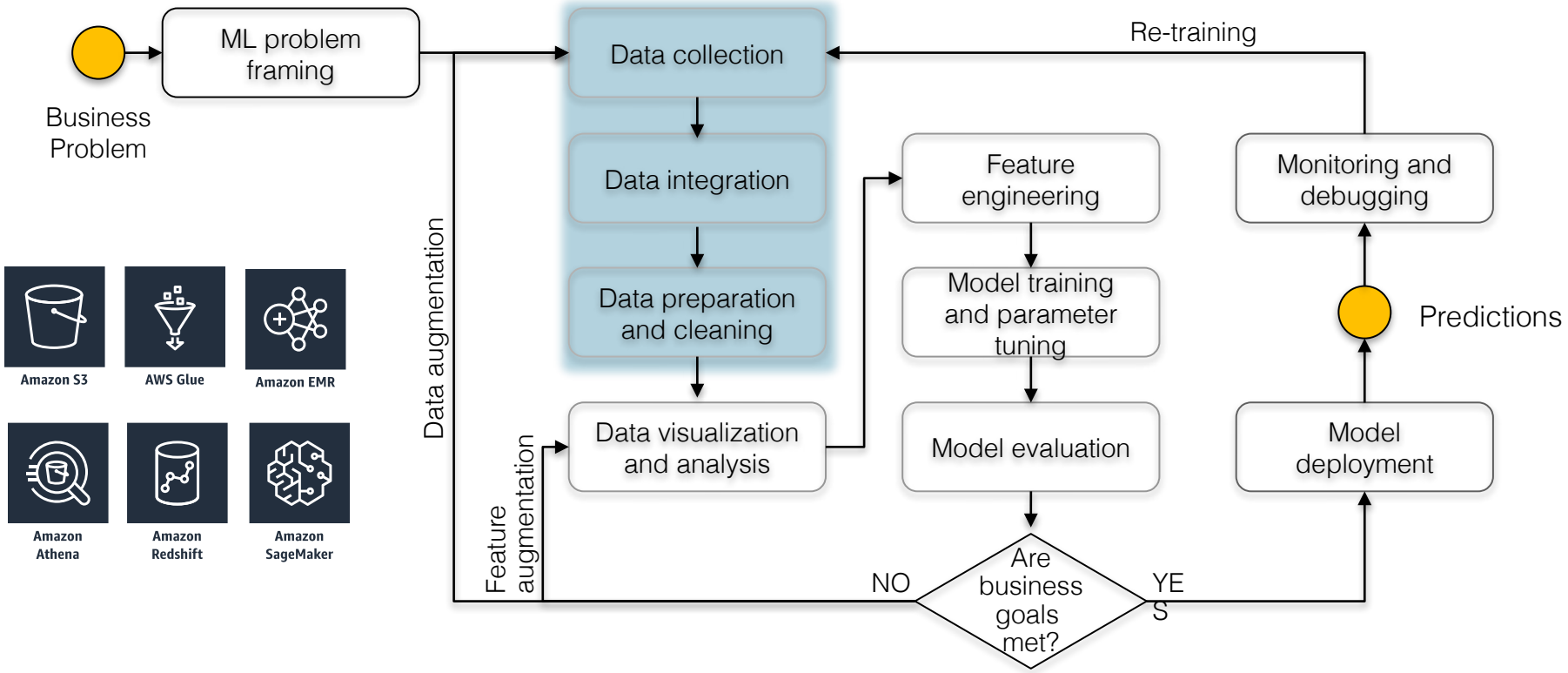
Automatic labeling via machine learning

Ready-made and custom workflows for image bounding box, segmentation, and text

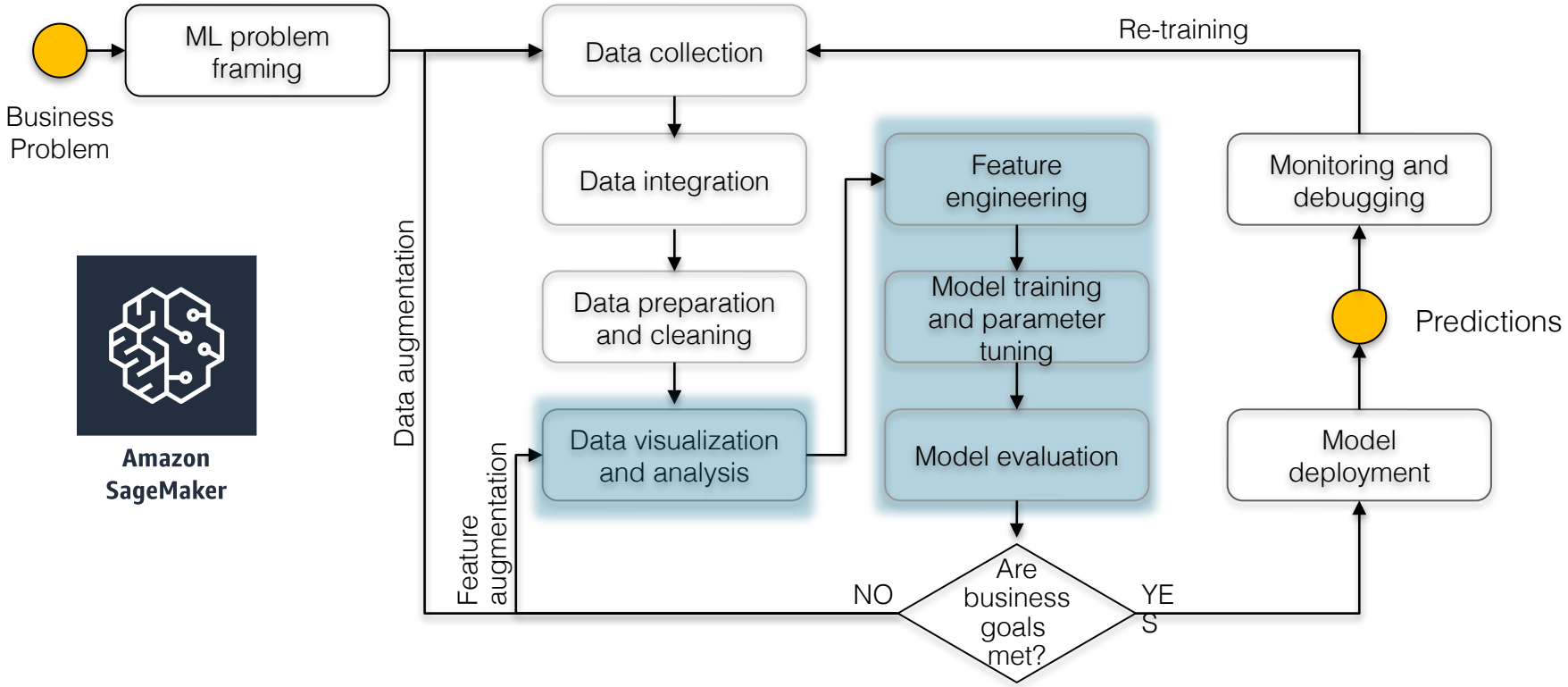
Private and public human workforce

Label management

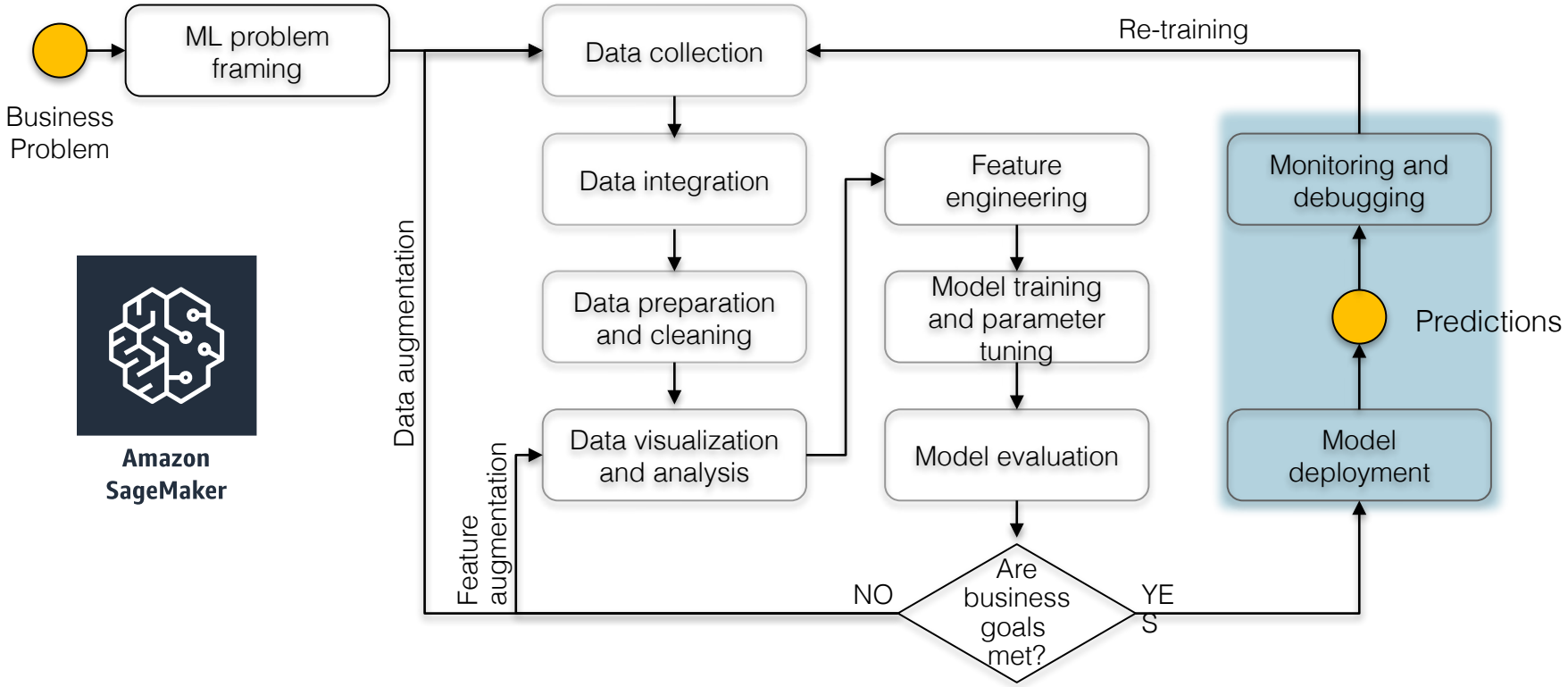
# Manage data on AWS



# Build and train models using SageMaker

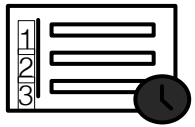


# Deploy models using SageMaker



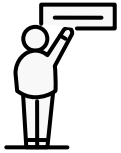
# Amazon SageMaker

Collect and prepare training data



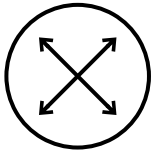

Choose and optimize your ML algorithm

Set up and manage environments for training



Train and Tune ML Models

Deploy models in production



Scale and manage the production environment

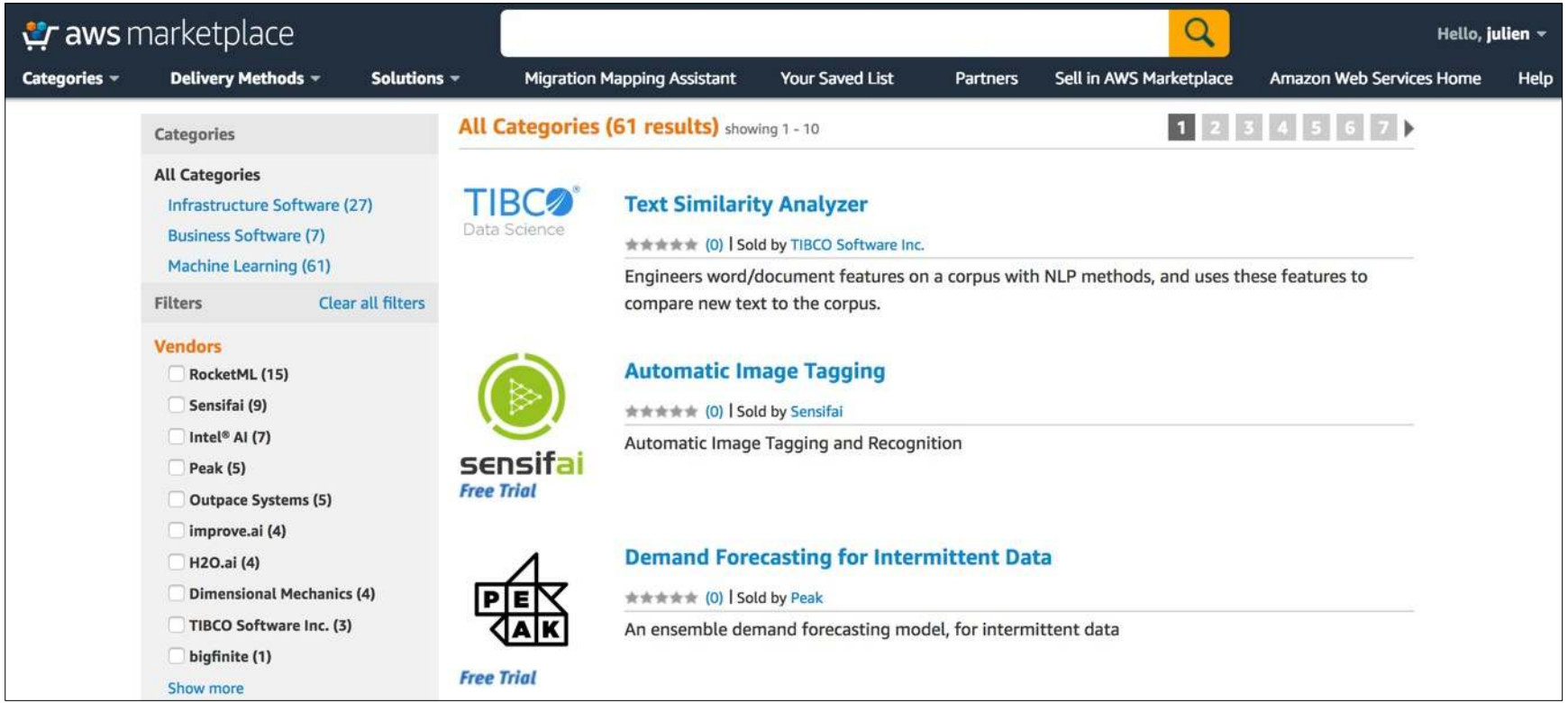
Git integration
Elastic inference

New built-in algorithms
scikit-learn environment
Model marketplace
Search

P3DN, C5N
TensorFlow on 256 GPUs
Resume HPO tuning job

Model compilation
Elastic inference
Inference pipelines

# Machine Learning Marketplace



DEV DAY

# Working with Amazon SageMaker

aws

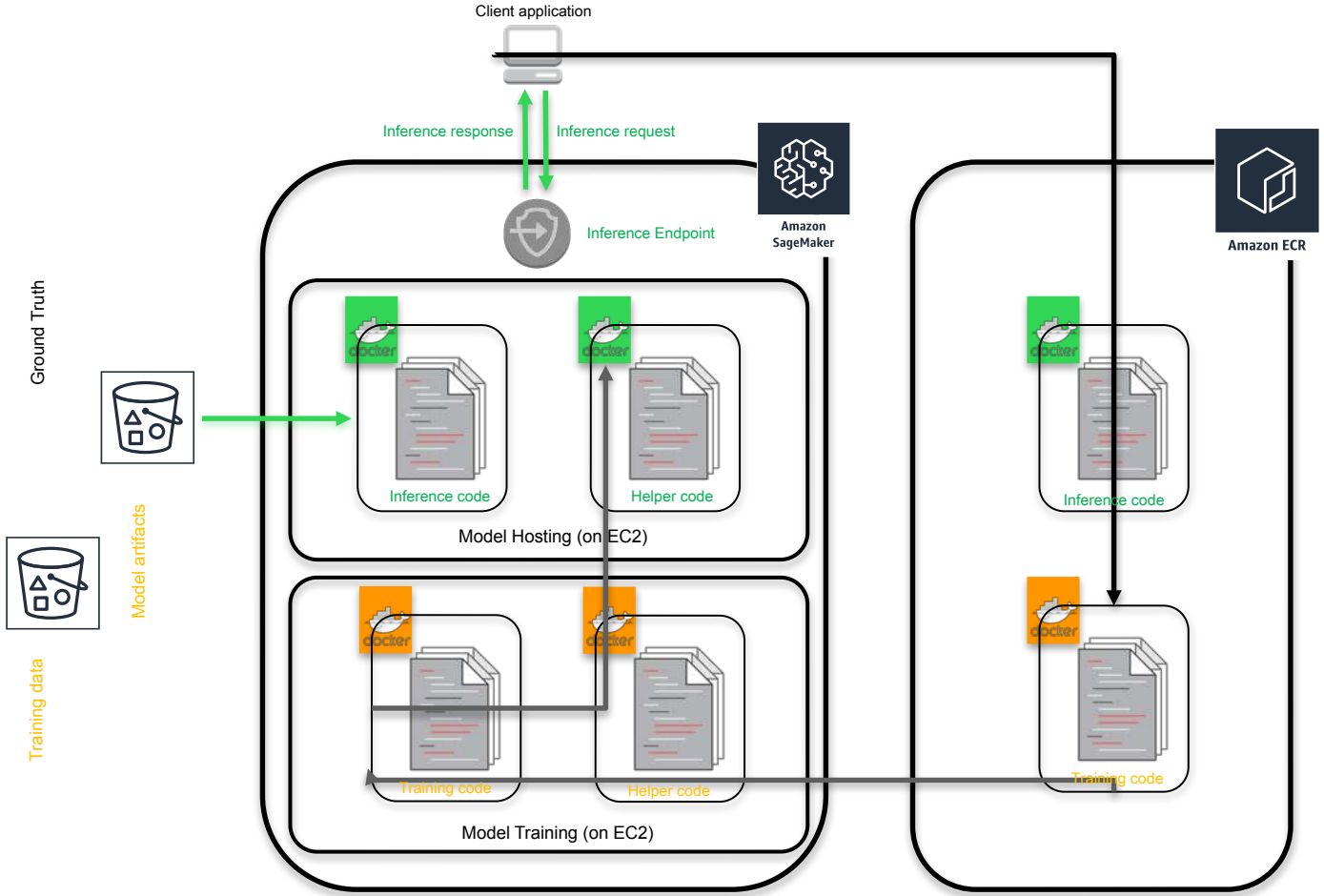
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# The Amazon SageMaker API

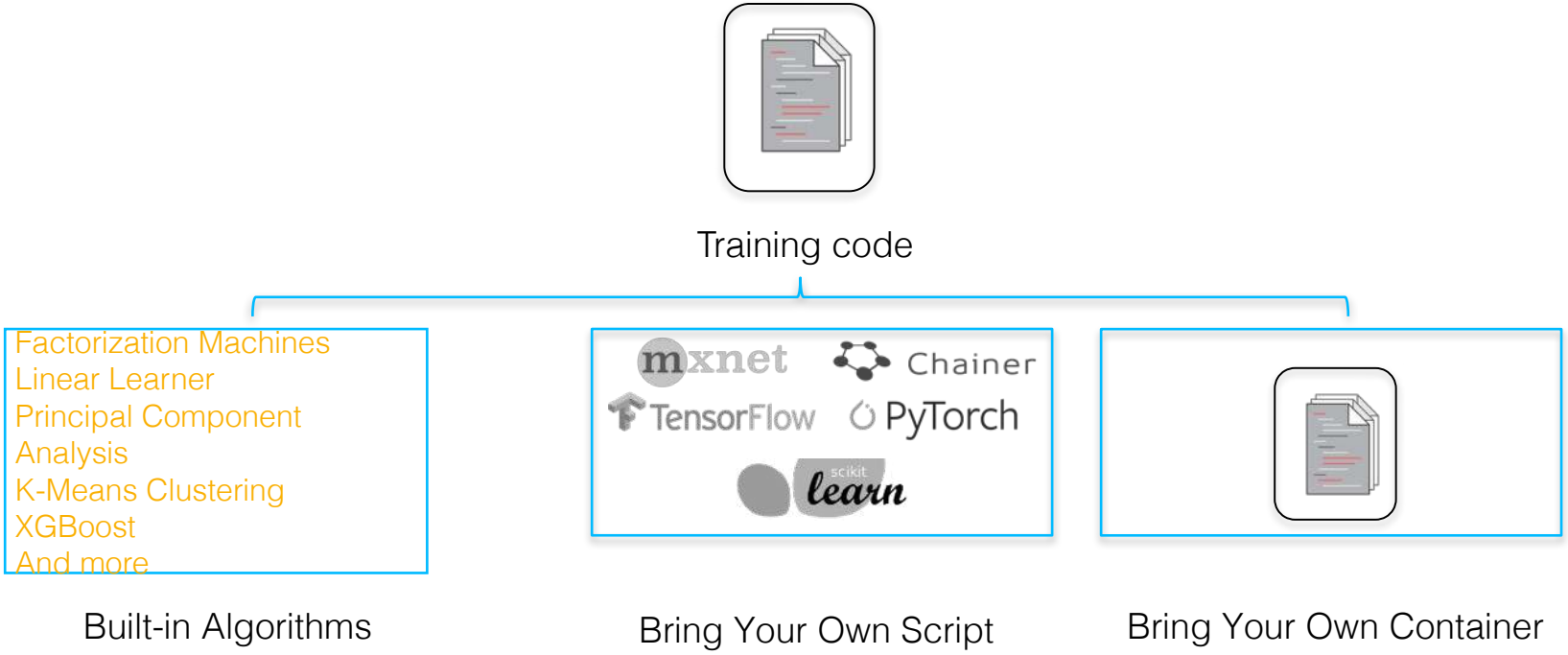
- Python SDK orchestrating all Amazon SageMaker activity
  - High-level objects for algorithm selection, training, deploying, automatic model tuning, etc.
  - Spark SDK (Python & Scala)

- AWS CLI: '*aws sagemaker*'

- AWS SDK: boto3, etc.


Client application
Inference response
Inference request
Inference Endpoint
Amazon SageMaker
Amazon ECR
Ground Truth
Model artifacts
Training data
Inference code
Helper code
Model Hosting (on EC2)
Training code
Helper code
Model Training (on EC2)
Inference code
Training code

# Model options



DEV DAY

# Built-in algorithms

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Built-in algorithms

orange: supervised, yellow: unsupervised

| | |
|---|---|
| **Linear Learner**: regression, classification | **Image Classification**: Deep Learning (ResNet) |
| **Factorization Machines**: regression, classification, recommendation | **Object Detection (SSD)**: Deep Learning (VGG or ResNet) |
| **K-Nearest Neighbors**: non-parametric regression and classification | **Neural Topic Model**: topic modeling |
| **XGBoost**: regression, classification, ranking https://github.com/dmlc/xgboost | **Latent Dirichlet Allocation**: topic modeling (mostly) |
| **K-Means**: clustering | **Blazing Text**: GPU-based Word2Vec, and text classification |
| **Principal Component Analysis**: dimensionality reduction | **Sequence to Sequence**: machine translation, speech to text and more |
| **Random Cut Forest**: anomaly detection | **DeepAR**: time-series forecasting (RNN) |
| **Object2Vec**: general-purpose embedding | **IP Insights**: usage patterns for IP addresses |
| **Semantic Segmentation**: Deep Learning | |

# Demo:
# Image classification with Caltech-256

https://gitlab.com/juliensimon/dlnotebooks/sagemaker/

# Blazing Text

**BlazingText: Scaling and Accelerating Word2Vec using Multiple GPUs**

Saurabh Gupta
Amazon Web Services
gsaur@amazon.com

Vineet Khare
Amazon Web Services
vkhare@amazon.com

https://dl.acm.org/citation.cfm?id=3146354

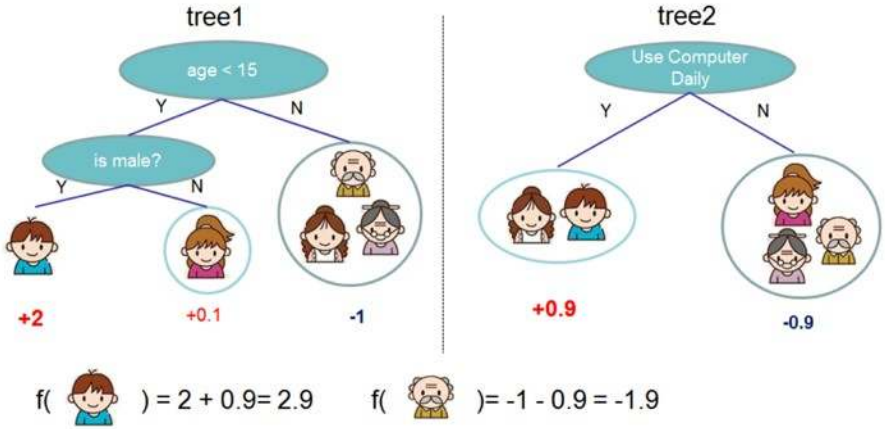
© 2017, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

# Demo:
# Text Classification with BlazingText

https://github.com/awslabs/amazon-sagemaker-examples/tree/master/introduction_to_amazon_algorithms/blazingtext_text_classification_dbpedia

# XGBoost

- Open Source project
- Popular tree-based algorithm for regression, classification and ranking
- Builds a collection of trees.
- Handles missing values and sparse data
- Supports distributed training
- Can work with data sets larger than RAM




https://github.com/dmlc/xgboost
https://xgboost.readthedocs.io/en/latest/
https://arxiv.org/abs/1603.02754

# Demo: XGBoost

AWS re:Invent 2018 workshop

https://gitlab.com/juliensimon/ent321

DEV DAY

# Built-in frameworks

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Demo:
# Keras/TensorFlow CNN on CIFAR-10

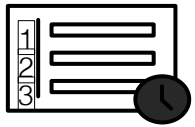
https://github.com/awslabs/amazon-sagemaker-examples/blob/master/sagemaker-python-sdk/tensorflow_keras_cifar10/tensorflow_keras_CIFAR10.ipynb

# Demo:
# Sentiment analysis with Apache MXNet

https://github.com/awslabs/amazon-sagemaker-examples/blob/master/sagemaker-python-sdk/mxnet_sentiment_analysis_with_gluon.ipynb

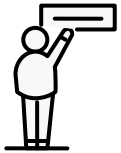
# Amazon SageMaker

Collect and prepare training data



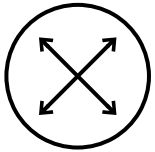

Choose and optimize your ML algorithm

Set up and manage environments for training



Train and Tune ML Models

Deploy models in production



Scale and manage the production environment

**Build**

**Train**

**Deploy**

FREE TIER

# Getting started

http://aws.amazon.com/free

https://ml.aws
https://aws.amazon.com/sagemaker

https://github.com/aws/sagemaker-python-sdk
https://github.com/aws/sagemaker-spark
https://github.com/awslabs/amazon-sagemaker-examples
https://gitlab.com/juliensimon/ent321

https://medium.com/@julsimon
https://gitlab.com/juliensimon/dlnotebooks

© 2018, Amazon Web Services, Inc. or Its Affiliates. All rights reserved.

DEV DAY

# Thank you!

Julien Simon
Global Evangelist, AI & Machine Learning
@julsimon
https://medium.com/@julsimon

aws

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.