



AWS  
re:Invent

**AIM401-R**

# Distributed training, tuning, and inference with TensorFlow in Amazon SageMaker

**Brent Rabowsky**

Machine Learning Specialist  
Amazon Web Services

**Rama Thamman**

R&D Manager  
Amazon Web Services

# Agenda

Amazon SageMaker overview

Q&A

Set up for workshop

Hands-on workshop modules

# Amazon SageMaker overview

# The AWS ML stack: Broadest and deepest set of capabilities

## AI services

Easily add intelligence to applications without machine learning skills

Vision | Documents | Speech | Language | Chatbots | Forecasting | Recommendations

---

## ML services

Build, train, and deploy machine learning models fast

Data labeling | Prebuilt algorithms and notebooks | One-click training and deployment

---

## ML frameworks and infrastructure











Flexibility and choice, highest-performing infrastructure

Support for ML frameworks | Compute options purpose-built for ML


# The AWS ML stack

## Broadest and deepest set of capabilities














### AI services

Vision			Speech		Language		Chatbots	Forecasting	Recommendations
 Amazon Rekognition Image	 Amazon Rekognition Video	 Amazon Textract	 Amazon Polly	 Amazon Transcribe	 Amazon Translate	 Amazon Comprehend and Amazon Comprehend Medical	 Amazon Lex	 Amazon Forecast	 Amazon Personalize

### ML services

 <b>Amazon SageMaker</b>	Ground Truth	Notebooks	Algorithms + marketplace	Reinforcement learning	Training	Optimization	Deployment	Hosting

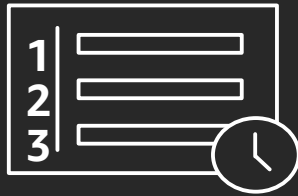
### ML frameworks + infrastructure

Frameworks	Interfaces	Infrastructure							
 TensorFlow  mxnet   PYTORCH	 GLUON   K Keras	 Amazon EC2 P3 and P3DN	 Amazon EC2 G4	 Amazon EC2 C5	 FPGAs	 DL containers and AMIs	 AWS IoT Greengrass	 Amazon Elastic Inference	 AWS Inferentia

# Amazon SageMaker: Build, train, and deploy ML models at scale



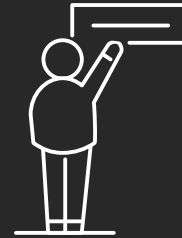
Collect and  
prepare training  
data



Choose and  
optimize your  
ML algorithm



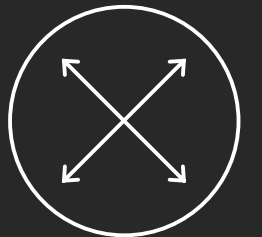
Set up and  
manage  
environments  
for training



Train and  
tune ML models



Deploy models  
in production

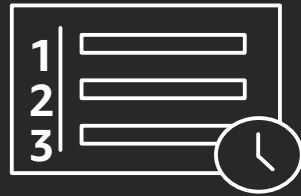


Scale and manage  
the production  
environment

# Amazon SageMaker: Build, train, and deploy ML models at scale

Prebuilt  
notebooks  
for common  
problems

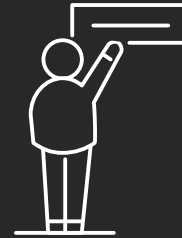
Collect and  
prepare training  
data



Choose and  
optimize your  
ML algorithm



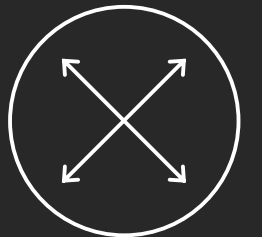
Set up and  
manage  
environments  
for training



Train and  
tune ML models



Deploy models  
in production



Scale and manage  
the production  
environment



# Amazon SageMaker: Build, train, and deploy ML models at scale

Prebuilt  
notebooks  
for common  
problems

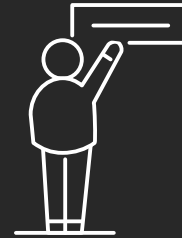
Collect and  
prepare training  
data

Built-in, high-  
performance  
algorithms

Choose and  
optimize your  
ML algorithm



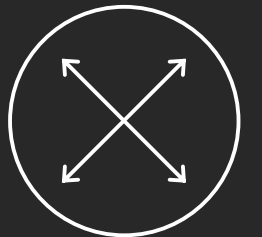
Set up and  
manage  
environments  
for training



Train and  
tune ML models



Deploy models  
in production



Scale and manage  
the production  
environment

# Amazon SageMaker: Build, train, and deploy ML models at scale

Prebuilt  
notebooks  
for common  
problems

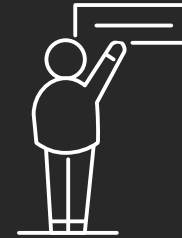
Collect and  
prepare training  
data

Built-in, high-  
performance  
algorithms

Choose and  
optimize your  
ML algorithm

One-click  
training on the  
highest-  
performing  
infrastructure

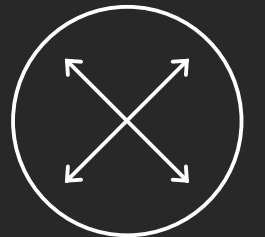
Set up and  
manage  
environments  
for training



Train and  
tune ML models



Deploy models  
in production



Scale and manage  
the production  
environment

# Amazon SageMaker: Build, train, and deploy ML models at scale

Prebuilt  
notebooks  
for common  
problems

Collect and  
prepare training  
data

Built-in, high-  
performance  
algorithms

Choose and  
optimize your  
ML algorithm

One-click  
training on the  
highest-  
performing  
infrastructure

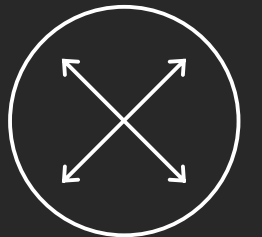
Set up and  
manage  
environments  
for training

Model  
optimization

Train and  
tune ML models



Deploy models  
in production



Scale and manage  
the production  
environment

# Amazon SageMaker: Build, train, and deploy ML models at scale

Prebuilt  
notebooks  
for common  
problems

Collect and  
prepare training  
data

Built-in, high-  
performance  
algorithms

Choose and  
optimize your  
ML algorithm

One-click  
training on the  
highest-  
performing  
infrastructure

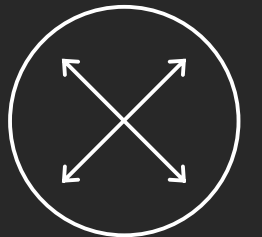
Set up and  
manage  
environments  
for training

Model  
optimization

Train and  
tune ML models

One-click  
deployment

Deploy models  
in production



Scale and manage  
the production  
environment

# Amazon SageMaker: Build, train, and deploy ML models at scale

Prebuilt notebooks for common problems

Collect and prepare training data

Built-in, high-performance algorithms

Choose and optimize your ML algorithm

One-click training on the highest-performing infrastructure

Set up and manage environments for training

Model optimization

Train and tune ML models

One-click deployment

Deploy models in production

Fully managed with AWS Auto Scaling for 75% less

Scale and manage the production environment

Intuit



CONVOY

SIEMENS



DOW JONES



Sony



# Q&A

# Set up for workshop

# Workshop options

**Lab 1 - NLP Use Case** – Sentiment analysis: Train a textCNN to classify movie reviews as positive or negative and perform batch inference on a set of reviews

- Local mode training
- Hosted training mode
- Batch prediction



# Workshop options, cont'd

**Lab 2 - Structured Data Use Case** – Boston housing: While training a regression MLP to predict house prices, use automatic model tuning to find the best hyperparameters

- Local mode training
- Local mode endpoint
- Hosted training mode with model extraction for running it elsewhere
- Automatic model tuning

# Workshop options, cont'd

**Lab 3 - Computer Vision Use Case** – Image classification: Train a CNN from scratch to classify images. This notebook showcases two distributed training approaches.

- Distributed training using parameter servers
- Distributed training using Horovod

# Workshop setup

<https://github.com/rthamman/reinvent-2019/>

Click on README.md

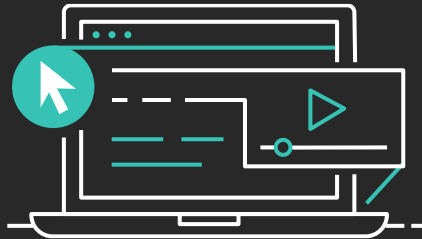
# Hands-on workshop modules

# Learn ML with AWS Training and Certification

The same training that our own developers use, now available on demand



Role-based ML learning paths for developers, data scientists, data platform engineers, and business decision makers



70+ free digital ML courses from AWS experts let you learn from real-world challenges tackled at AWS



Validate expertise with the  
**AWS Certified Machine Learning - Specialty** exam

Visit <https://aws.training/machinelearning>

# Thank you!

**Brent Rabowsky**

rabowsky@amazon.com

**Rama Thamman**

rathamma@amazon.com



Please complete the session  
survey in the mobile app.