



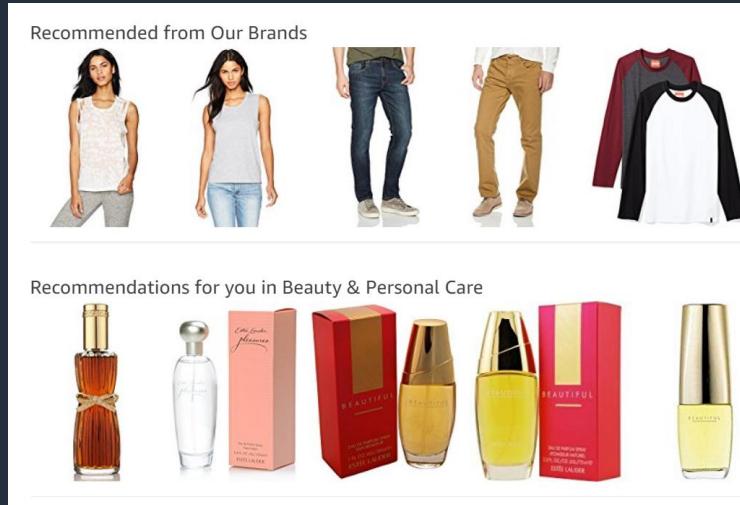
# INNOVATE2018

ONLINE CONFERENCE



## Manage Deep Learning Models at Scale with Amazon SageMaker (Level 300)

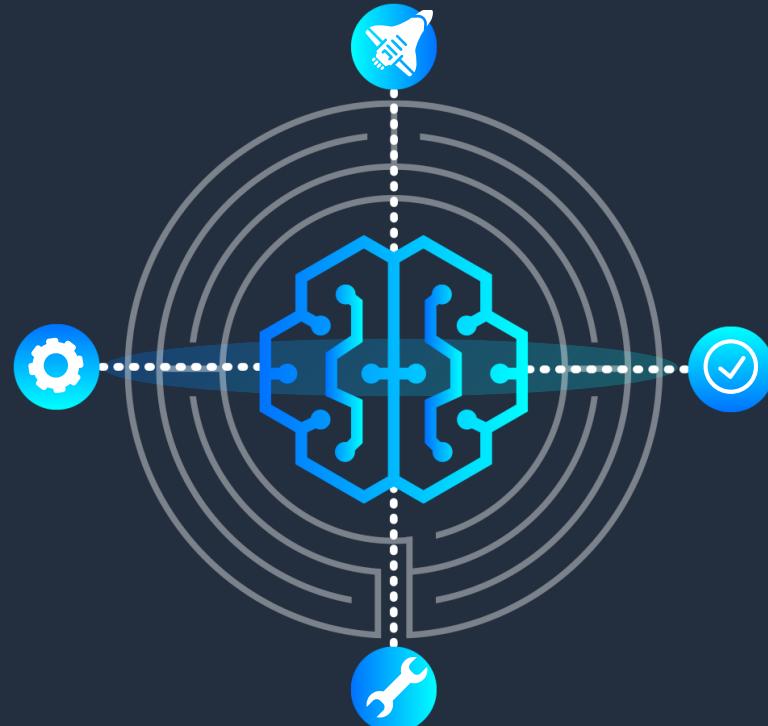
Jan Haak, Global Solution Architect



20+ Years of  
**Machine Learning**  
Innovation  
@

amazon





# Amazon SageMaker

Easily build, train, and deploy  
machine learning models

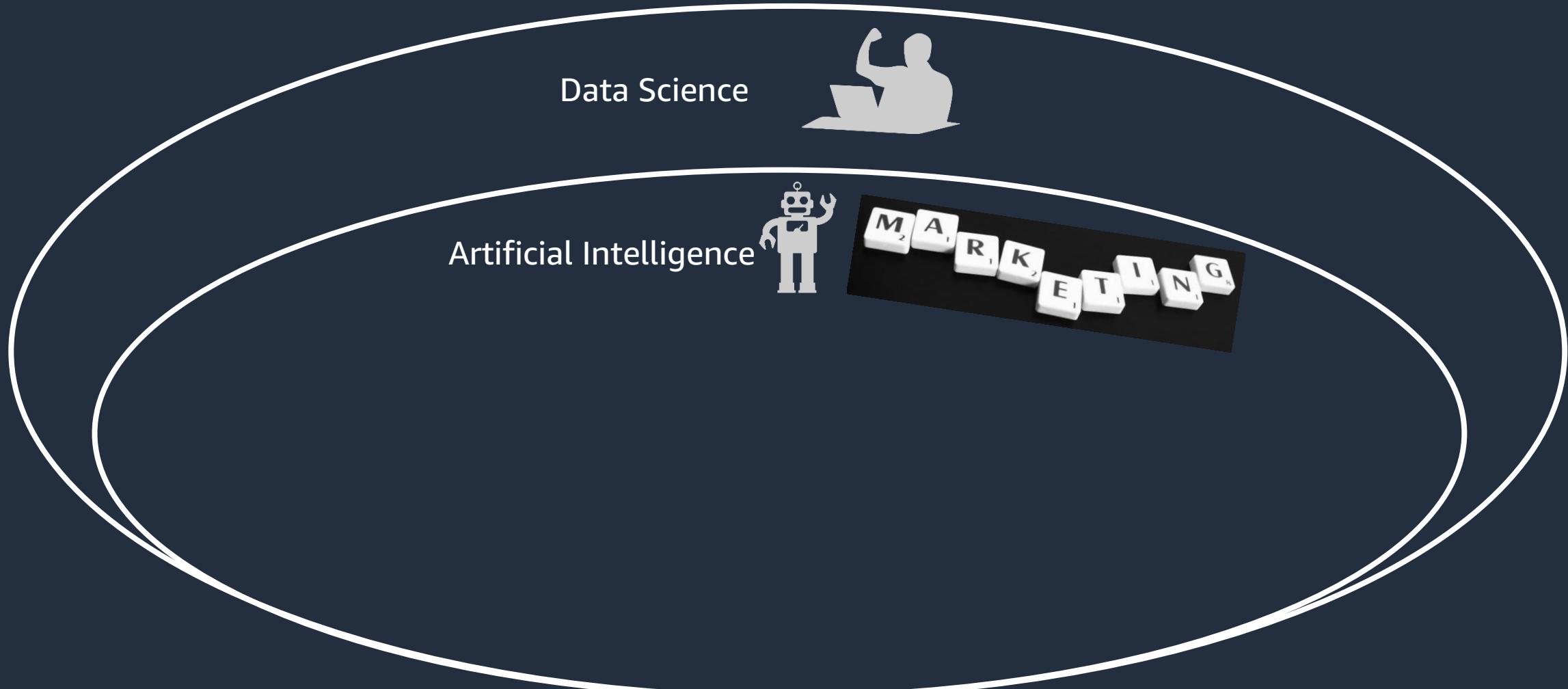
# Data Science $\nu s$ AI $\nu s$ Machine Learning $\nu s$ Deep Learning

# Data Science *vs* AI *vs* Machine Learning *vs* Deep Learning

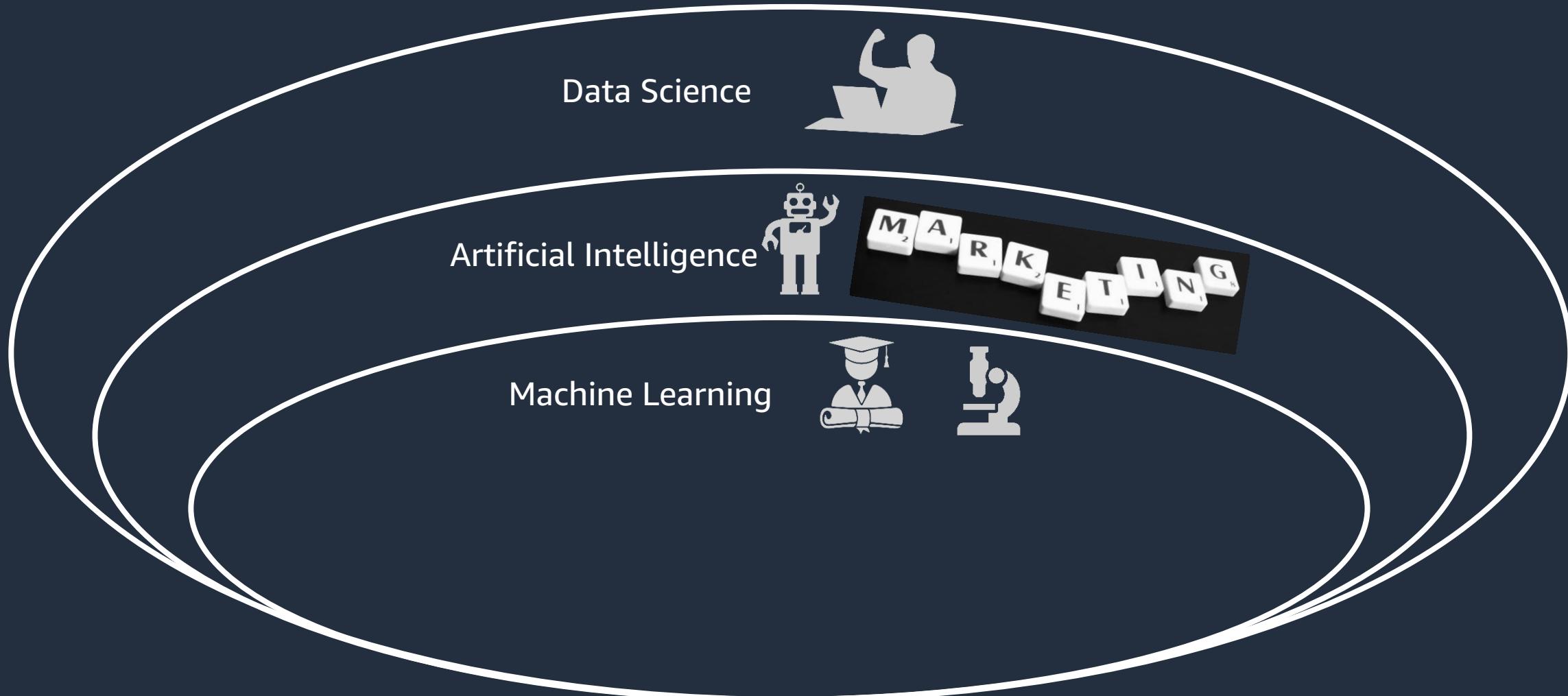
Data Science



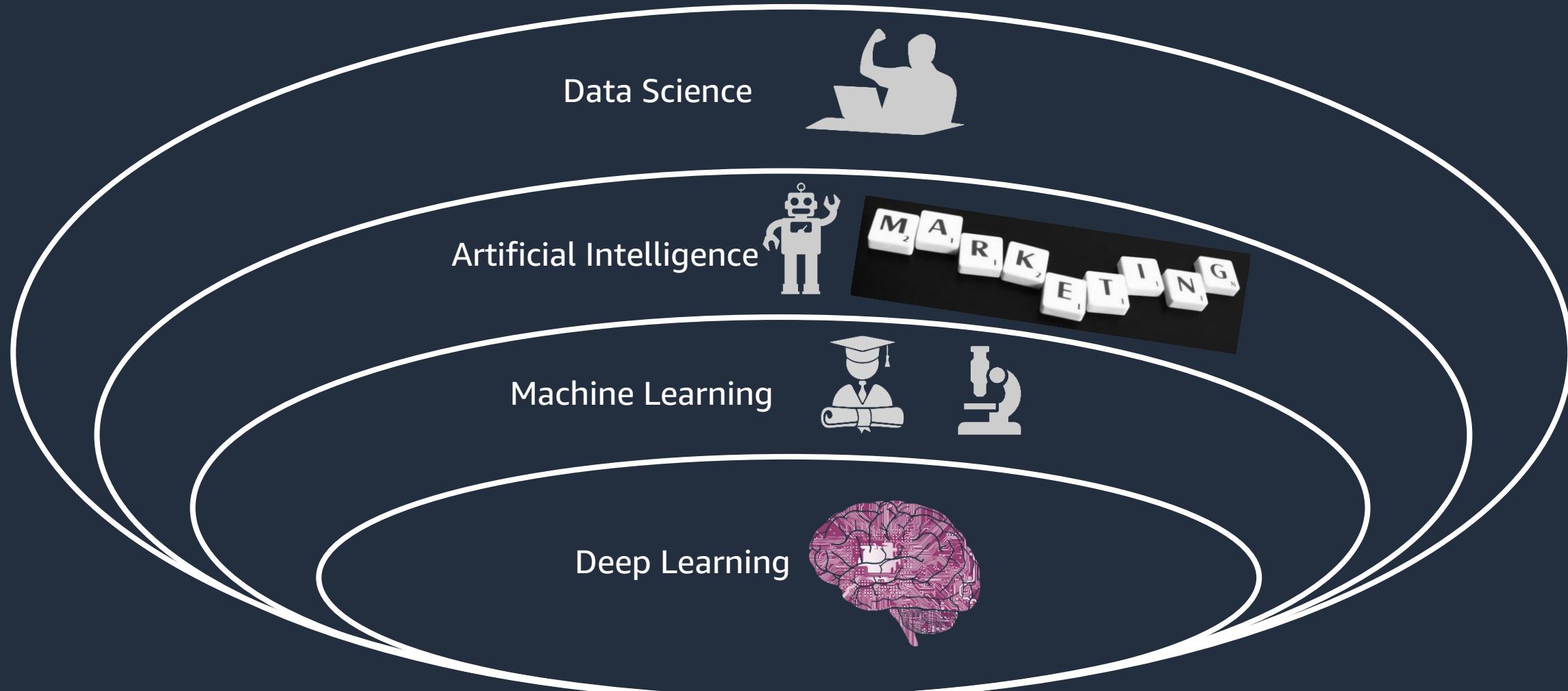
# Data Science *vs* AI *vs* Machine Learning *vs* Deep Learning



# Data Science *vs* AI *vs* Machine Learning *vs* Deep Learning



# Data Science *vs* AI *vs* Machine Learning *vs* Deep Learning



# Haven't I Heard All This Before? Why Now?

# Haven't I Heard All This Before? Why Now?

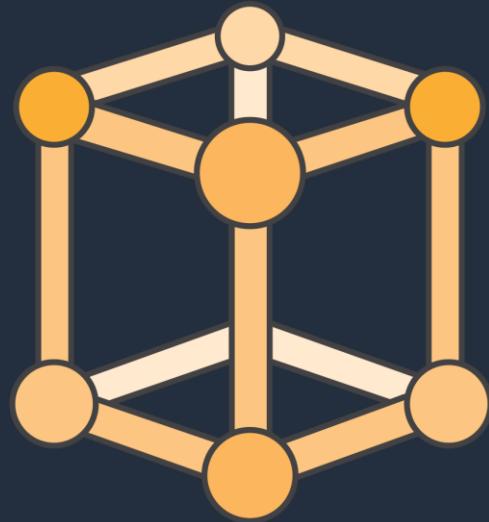


Vast data

# Haven't I Heard All This Before? Why Now?



Vast data

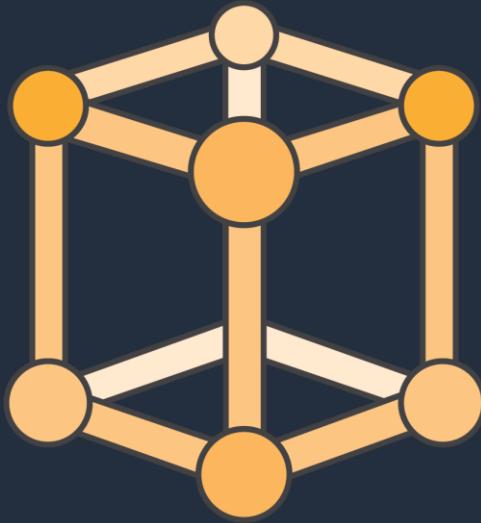


Algorithms

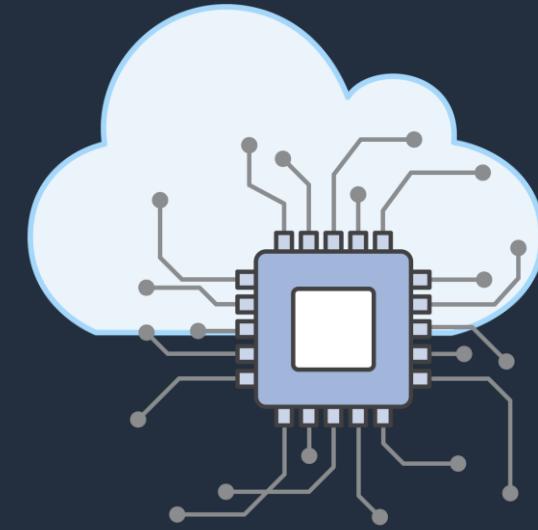
# Haven't I Heard All This Before? Why Now?



Vast data



Algorithms



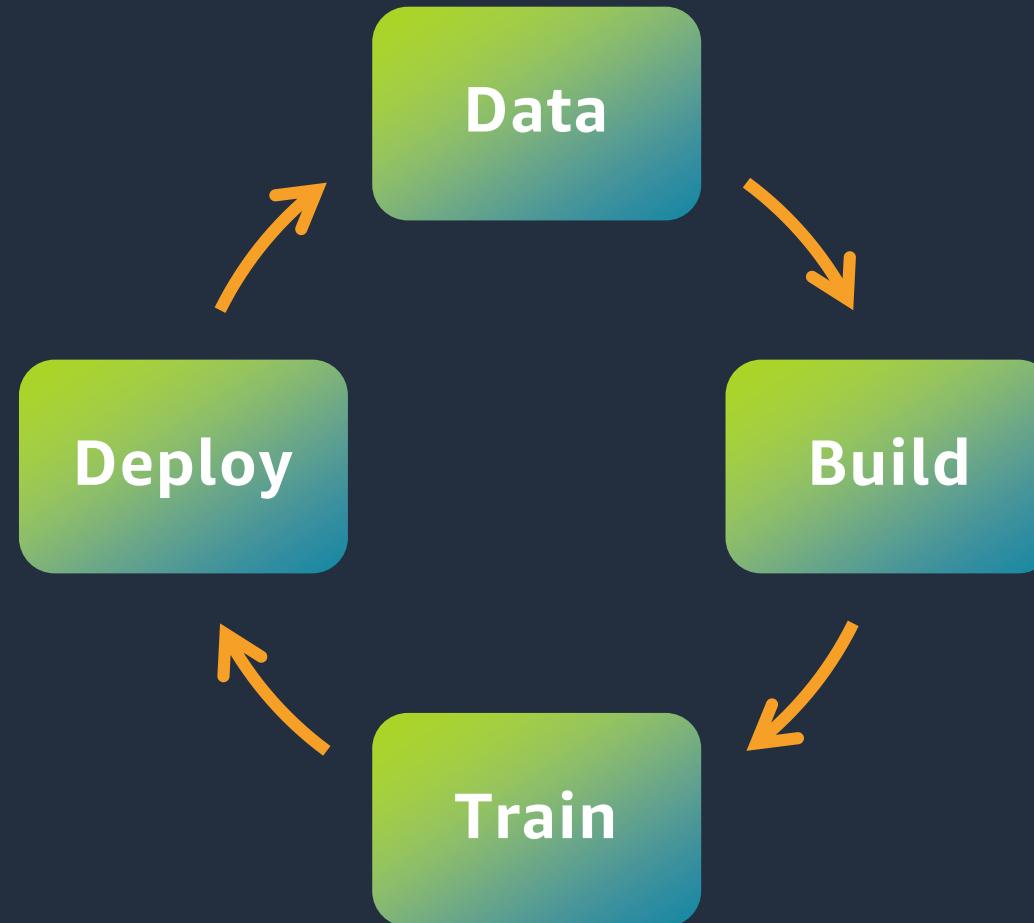
Compute

# ML @ AWS: Our mission

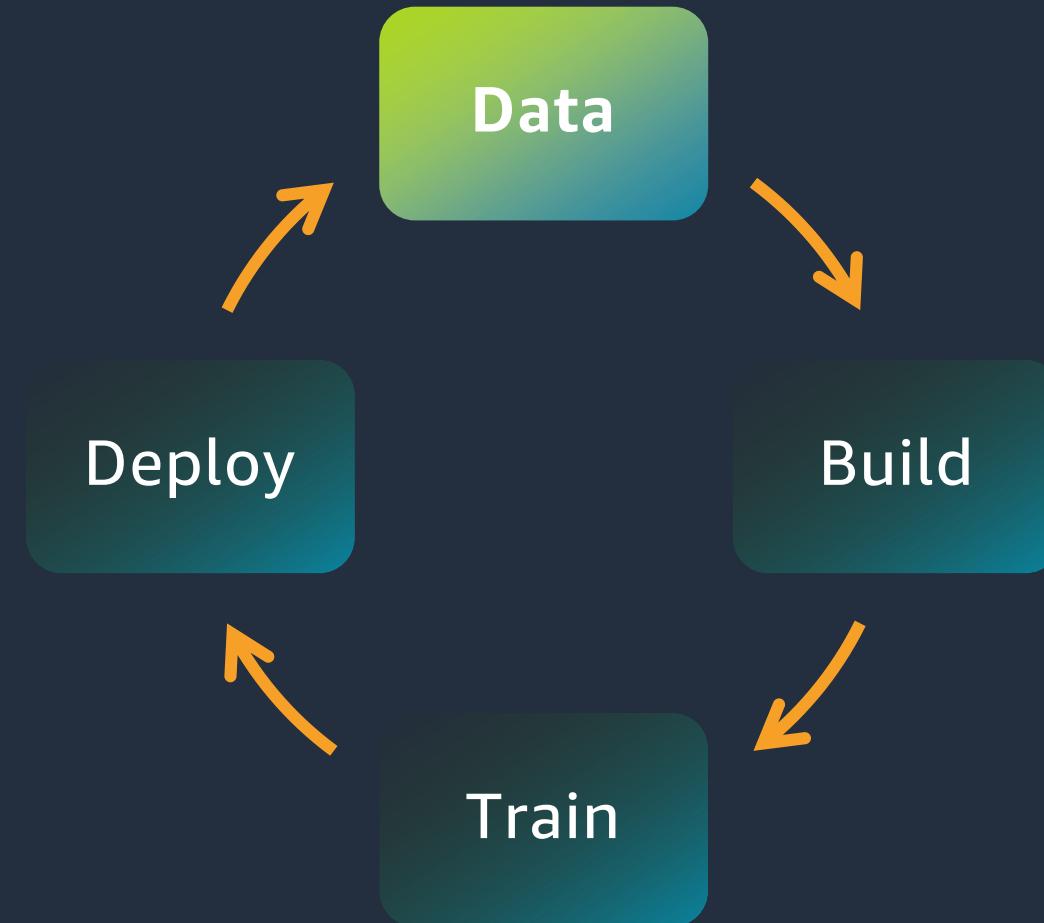
Put machine learning in the hands of every developer  
and data scientist



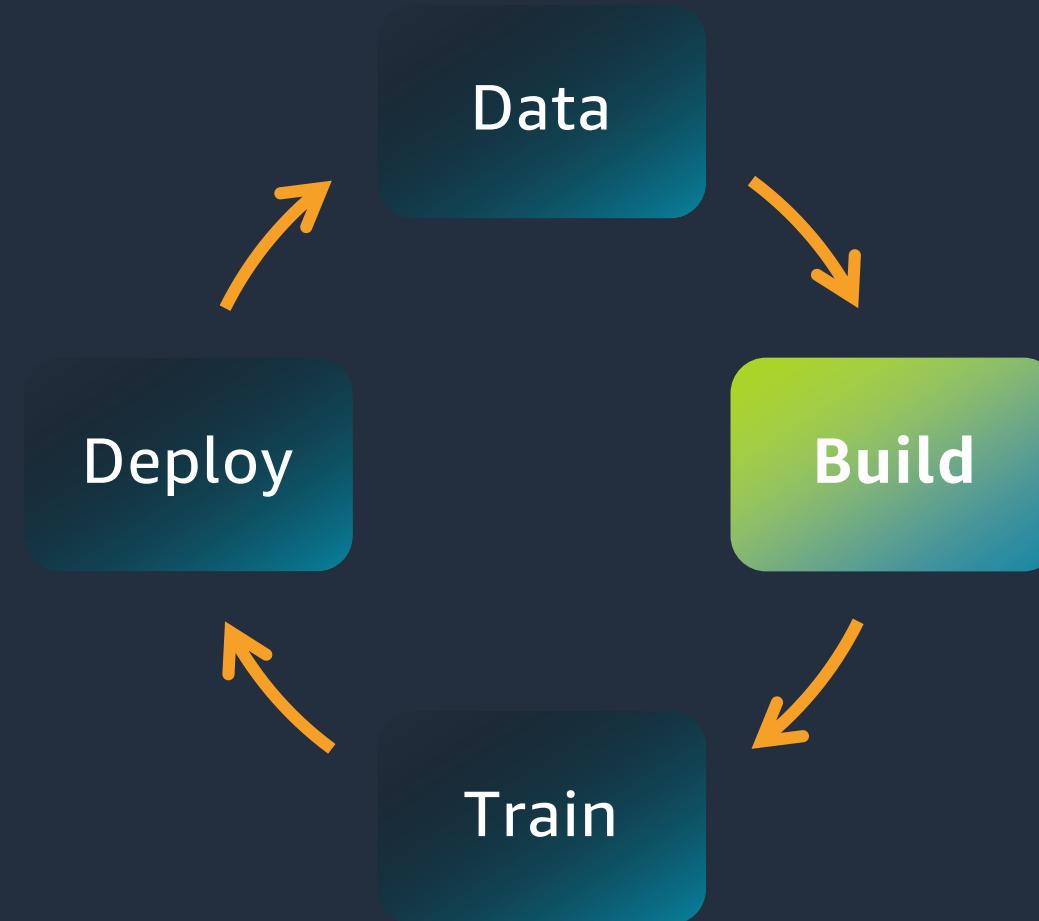
# A Typical Machine Learning Pipeline



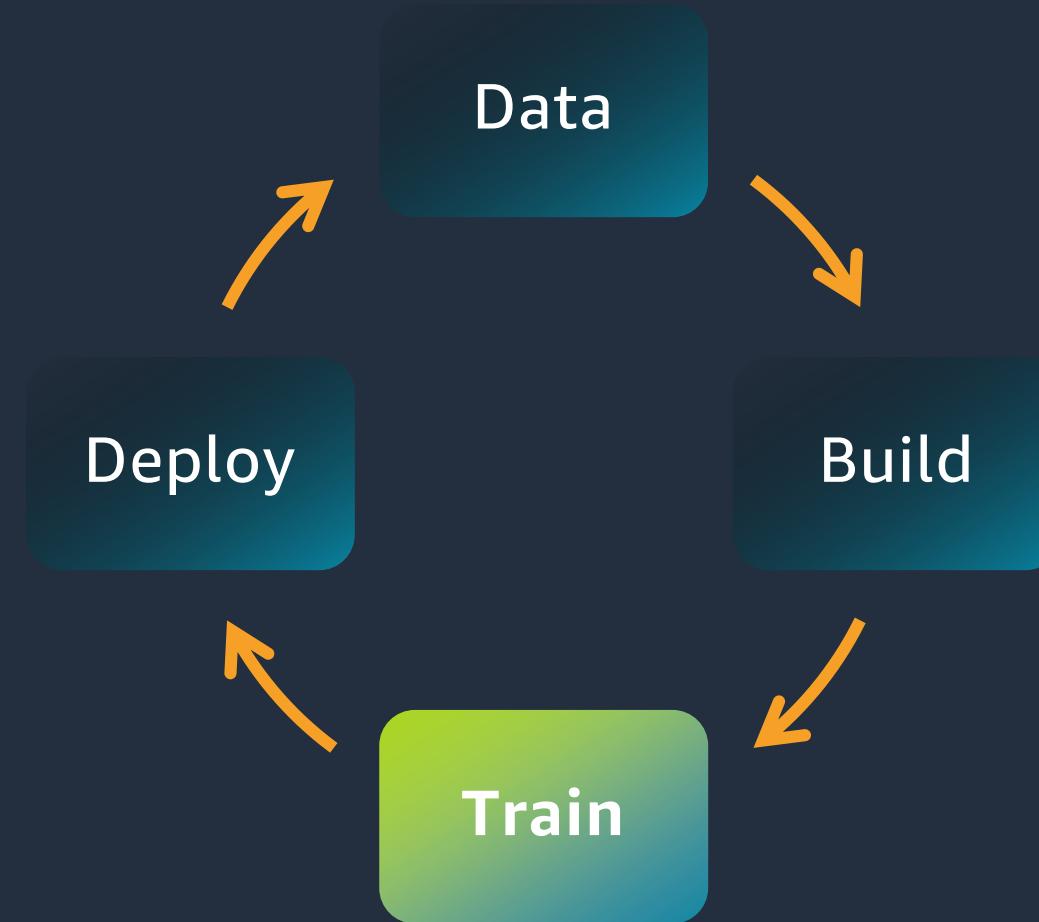
# A Typical Machine Learning Pipeline



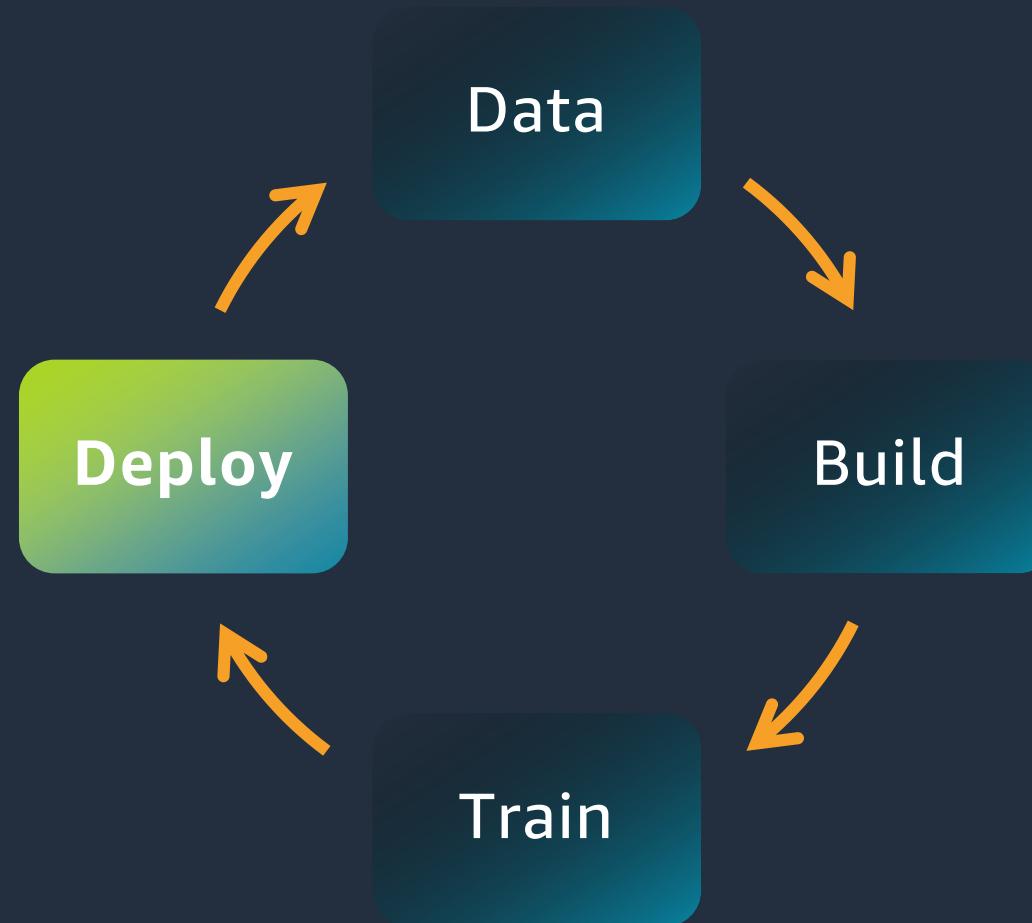
# A Typical Machine Learning Pipeline



# A Typical Machine Learning Pipeline

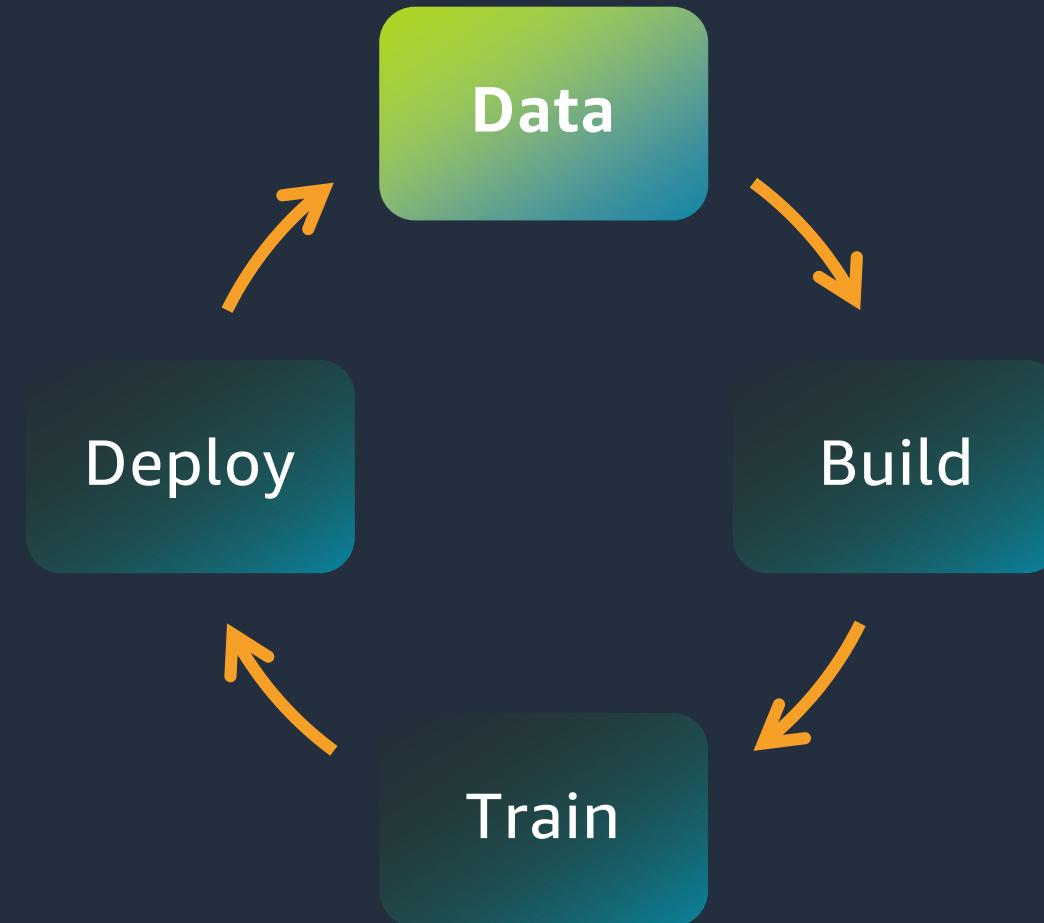


# A Typical Machine Learning Pipeline





# A Typical Machine Learning Pipeline

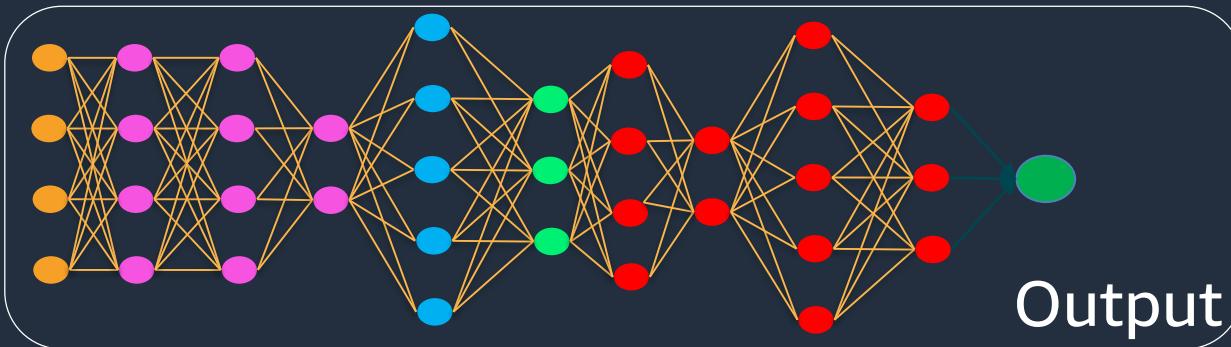


# Challenge: Lack Of Training Data

- Solution 1a: Transfer Learning

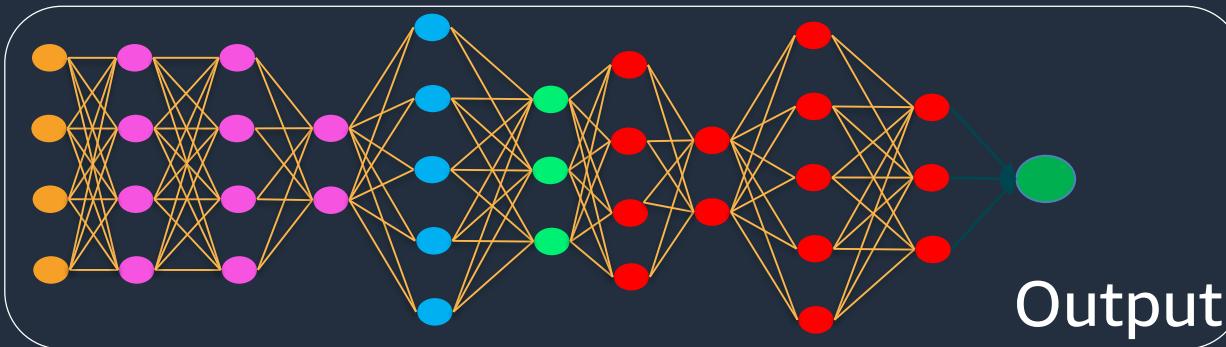
# Challenge: Lack Of Training Data

- Solution 1a: Transfer Learning

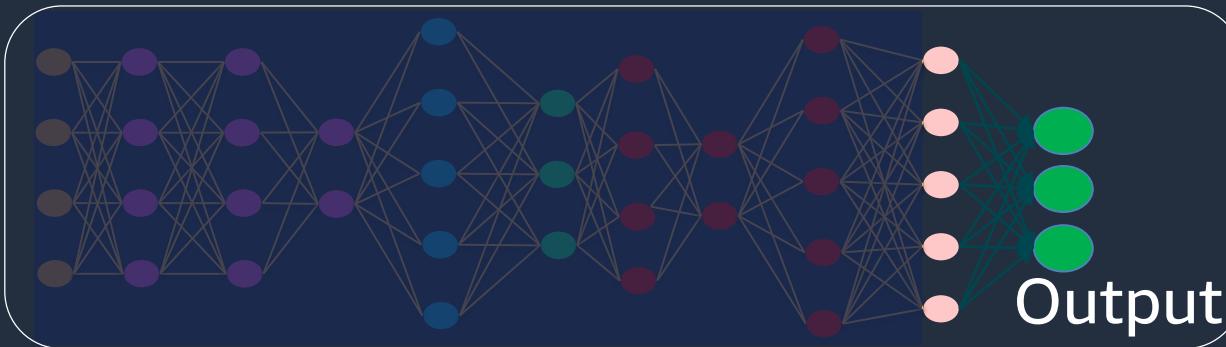


# Challenge: Lack Of Training Data

- Solution 1a: Transfer Learning



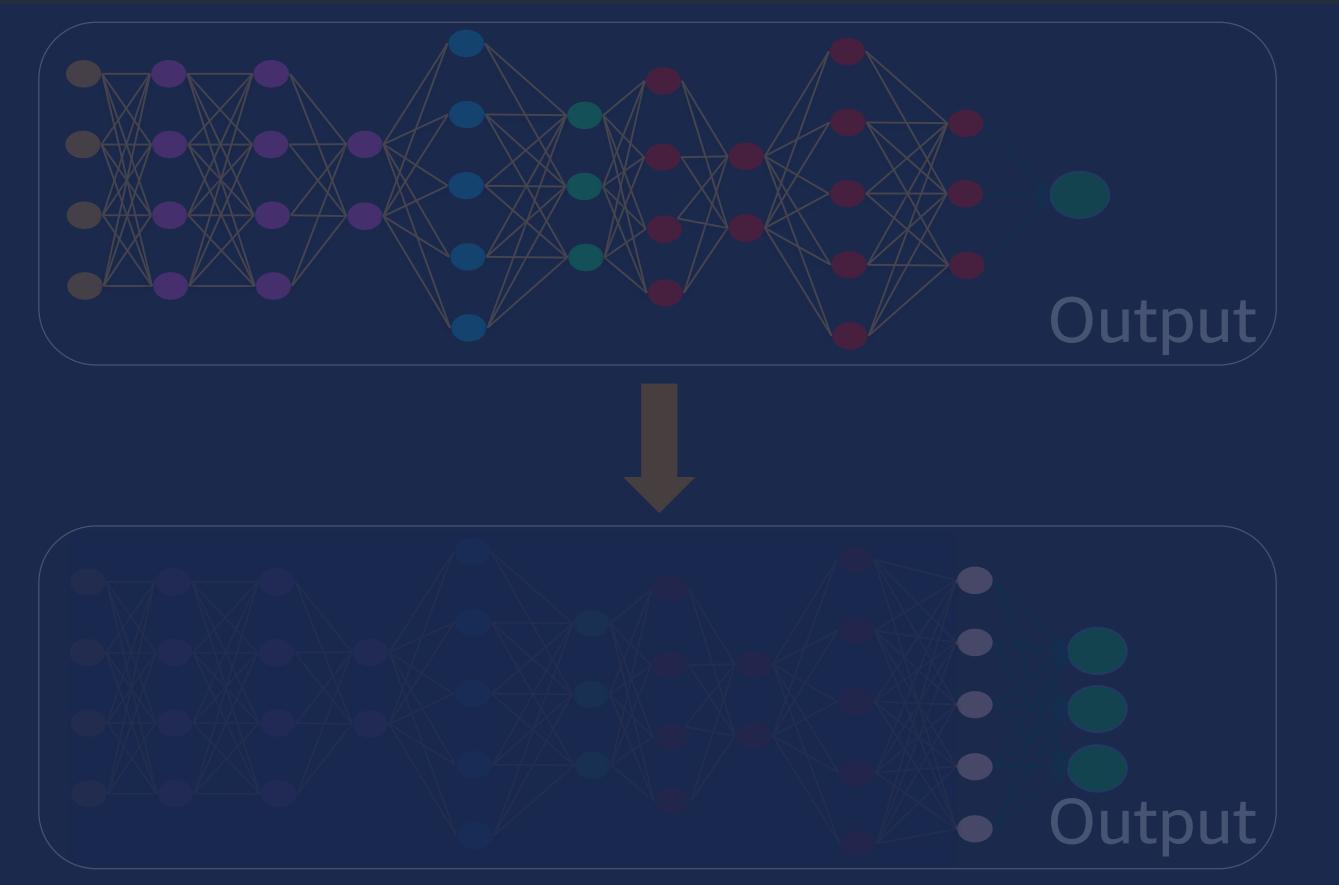
Output



Output

# Challenge: Lack Of Training Data

- Solution 1a: Transfer Learning



- Solution 1b: Public Datasets

- ImageNet
- MNIST
- CIFAR
- Caltech256



# Building Models - The New Way



Notebooks



Amazon EC2



Deep  
Learning AMI



Pandas



TensorFlow™



scikit  
learn



NumPy



jupyter



Flask



mxnet



matplotlib



APACHE  
Spark™

---

BUILD

# Building Models - The New Way



Notebooks



Amazon EC2



Deep  
Learning AMI



Pandas



TensorFlow™



scikit-learn



NumPy



jupyter



Flask



mxnet



matplotlib



APACHE  
Spark™



Amazon SageMaker Notebook  
instance

BUILD



Services ▾

Resource Groups ▾



admin @ djenny ▾

Ohio ▾

Support ▾

Amazon SageMaker X

ARTIFICIAL INTELLIGENCE

# Amazon SageMaker

## Build, train, and deploy machine learning models at scale

The quickest and easiest way to get ML models from idea to production.

### Get started

Explore AWS data in your notebooks, and use algorithms to create models via training jobs. Leverage Notebook instances in the cloud to begin.

[Create notebook instance](#)[Start with an overview](#)

### Pricing (US)

With Amazon SageMaker, you pay only for what you use. Authoring, training and hosting is billed by the second, with no

### How it works

[Feedback](#)[English \(US\)](#)



Services ▾

Resource Groups ▾



admin @ djenny ▾

Ohio ▾

Support ▾

Amazon SageMaker X

ARTIFICIAL INTELLIGENCE

# Amazon SageMaker

## Build, train, and deploy machine learning models at scale

The quickest and easiest way to get ML models from idea to production.

### How it works

#### Get started

Explore AWS data in your notebooks, and use algorithms to create models via training jobs. Leverage Notebook instances in the cloud to begin.

[Create notebook instance](#)[Start with an overview](#)

#### Pricing (US)

With Amazon SageMaker, you pay only for what you use. Authoring, training and hosting is billed by the second, with no

[Feedback](#)[English \(US\)](#)



Services ▾

Resource Groups ▾



admin @ djenny

Ohio ▾

Support ▾

Include example code for common model training and hosting exercises. [Learn More](#)

## Amazon SageMaker X

Dashboard

Notebook instances

Jobs

### Resources

Models

Endpoint configuration

Endpoints

### Notebook instance settings

Notebook instance name

Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Notebook instance type

IAM role

Notebook instances require permissions to call other services including SageMaker and S3. Choose a role or let us create a role with the [AmazonSageMakerFullAccess](#) IAM policy attached.VPC - *optional*

Notebook instances will have internet access independent of your VPC setting.

Encryption key - *optional*

Encrypt your notebook data. Choose an existing KMS key or enter a key's ARN.

### ► Tags - *optional*

Cancel

Create notebook instance



Services ▾

Resource Groups ▾



admin @ djenny ▾

Ohio ▾

Support ▾

Include example code for common model training and hosting exercises. [Learn More](#)

## Amazon SageMaker X

Dashboard

Notebook instances

Jobs

### Resources

Models

Endpoint configuration

Endpoints

### Notebook instance settings

Notebook instance name

myFirstNotebook

Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Notebook instance type

ml.t2.medium

IAM role

Notebook instances require permissions to call other services including SageMaker and S3. Choose a role or let us create a role with the [AmazonSageMakerFullAccess](#) IAM policy attached.

AmazonSageMaker-ExecutionRole-20171214T162069

VPC - *optional*

Notebook instances will have internet access independent of your VPC setting.

No VPC

Encryption key - *optional*

Encrypt your notebook data. Choose an existing KMS key or enter a key's ARN.

No Encryption

### ► Tags - *optional*

Cancel

Create notebook instance



Services ▾

Resource Groups ▾



admin @ djenny ▾

Ohio ▾

Support ▾

Include example code for common model training and hosting exercises. [Learn More](#)

## Amazon SageMaker X

Dashboard

Notebook instances

Jobs

### Resources

Models

Endpoint configuration

Endpoints

### Notebook instance settings

Notebook instance name

myFirstNotebook

Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Notebook instance type

ml.t2.medium

IAM role

Notebook instances require permissions to call other services including SageMaker and S3. Choose a role or let us create a role with the [AmazonSageMakerFullAccess](#) IAM policy attached.

AmazonSageMaker-ExecutionRole-20171214T162069

VPC - optional

Notebook instances will have internet access independent of your VPC setting.

No VPC

Encryption key - optional

Encrypt your notebook data. Choose an existing KMS key or enter a key's ARN.

No Encryption

### ► Tags - optional

Cancel

Create notebook instance



Services ▾

Resource Groups ▾



admin @ djenny ▾

Ohio ▾

Support ▾

## Amazon SageMaker X

Dashboard

Notebook instances

Jobs

## Resources

Models

Endpoint configuration

Endpoints

Success! Your notebook instance is being created.

Open the notebook instance when status is InService and open a template notebook to get started.

[View details](#)[Amazon SageMaker](#) > Notebook instances

### Notebook instances

[Open](#)[Start](#)[Update settings](#)[Actions ▾](#)[Create notebook instance](#) Search notebook instances< 1 > [⚙️](#)

Name	Instance	Creation time	Status	Actions
<a href="#">myFirstNotebook</a>	ml.t2.medium	Mar 10, 2018 03:44 UTC	Pending	—

[Feedback](#) [English \(US\)](#)



Services ▾

Resource Groups ▾



admin @ djenny ▾

Ohio ▾

Support ▾

## Amazon SageMaker X

Dashboard

Notebook instances

Jobs

### Resources

Models

Endpoint configuration

Endpoints

Success! Your notebook instance is being created.

Open the notebook instance when status is InService and open a template notebook to get started.

[View details](#)[Amazon SageMaker](#) > Notebook instances

### Notebook instances

[Open](#)[Start](#)[Update settings](#)[Actions ▾](#)[Create notebook instance](#) Search notebook instances< 1 >

Name	Instance	Creation time	Status	Actions
myFirstNotebook	ml.t2.medium	Mar 10, 2018 03:44 UTC	InService	<a href="#">Open   Stop</a>



Services ▾

Resource Groups ▾



admin @ djenny ▾

Ohio ▾

Support ▾

## Amazon SageMaker X

Dashboard

Notebook instances

Jobs

### Resources

Models

Endpoint configuration

Endpoints

Success! Your notebook instance is being created.

Open the notebook instance when status is InService and open a template notebook to get started.

[View details](#)[Amazon SageMaker](#) > Notebook instances

### Notebook instances

[Open](#)[Start](#)[Update settings](#)[Actions ▾](#)[Create notebook instance](#)

Search notebook instances

&lt; 1 &gt;



Name	Instance	Creation time	Status	Actions
myFirstNotebook	ml.t2.medium	Mar 10, 2018 03:44 UTC	InService	<a href="#">Open   Stop</a>



Files    Running    Clusters    Conda

Select items to perform actions on them.

Upload    New ▾    ⚡

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Name ↑	Last Modified ↑
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	lost+found	6 minutes ago
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	sample-notebooks	2 months ago

# Amazon SageMaker Algorithms And Frameworks



Notebooks



Selecting  
Algorithms and  
Frameworks

## ALGORITHMS

K-Means Clustering  
Principal Component Analysis  
Neural Topic Modelling  
Factorisation Machines  
Linear Learner – Regression  
DeepAR Forecasting

XGBoost  
Latent Dirichlet Allocation  
Image Classification  
Seq2Seq  
Linear Learner – Classification

## FRAMEWORKS

Apache MXNet  
TensorFlow

Caffe2, CNTK,  
PyTorch, Torch

BUILD

# Training Models – The Painful Way

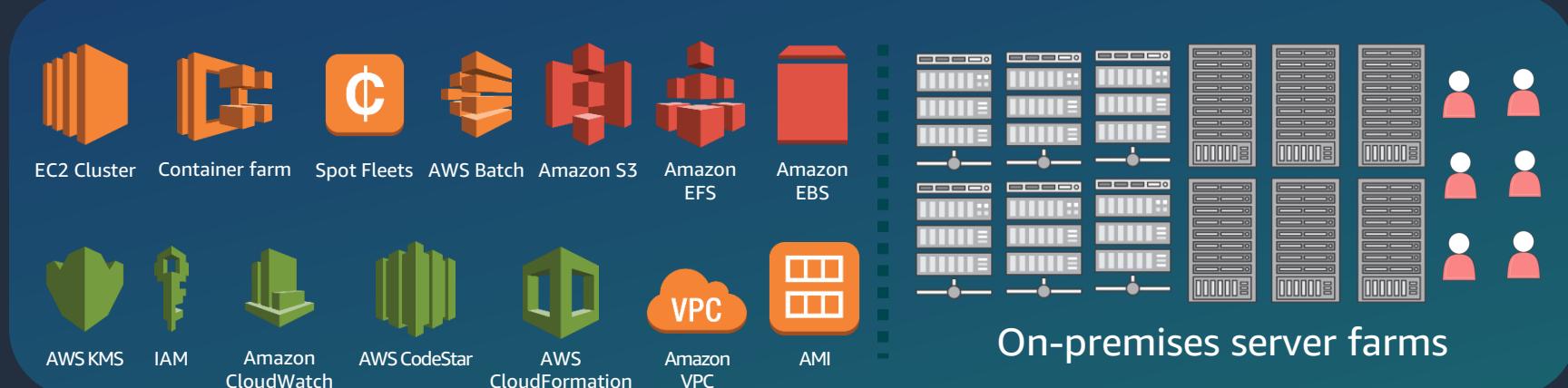


TRAIN

# Training Models – The Painful Way



Model  
Training



One-click Training and HPO

TRAIN



Files    Running    Clusters    Conda

Select items to perform actions on them.

Upload    New ▾    ⌂

<input type="checkbox"/>	<input type="checkbox"/>	Name ↑	Last Modified ↑
<input type="checkbox"/>	📁 lost+found		6 minutes ago
<input type="checkbox"/>	📁 sample-notebooks		2 months ago



Files    Running    Clusters    Conda

Select items to perform actions on them.

Upload    New ▾   

<input type="checkbox"/>	Name	Last Modified
<input type="checkbox"/>	sample-notebooks	
<input type="checkbox"/>	..	seconds ago
<input type="checkbox"/>	advanced_functionality	2 months ago
<input type="checkbox"/>	introduction_to_amazon_algorithms	2 months ago
<input type="checkbox"/>	introduction_to_applying_machine_learning	2 months ago
<input type="checkbox"/>	sagemaker-python-sdk	2 months ago
<input type="checkbox"/>	scientific_details_of_algorithms	2 months ago
<input type="checkbox"/>	LICENSE.txt	2 months ago
<input type="checkbox"/>	README.md	2 months ago



Files    Running    Clusters    Conda

Select items to perform actions on them.

Upload    New ▾   

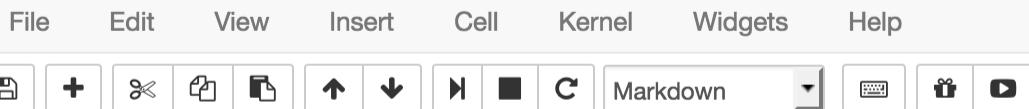
<input type="checkbox"/>	Name ↑	Last Modified ↑
<input type="checkbox"/>	..	seconds ago
<input checked="" type="checkbox"/>	1P_kmeans_highlevel	a minute ago
<input type="checkbox"/>	1P_kmeans_lowlevel	2 months ago
<input type="checkbox"/>	mxnet_gluon_cifar10	2 months ago
<input type="checkbox"/>	mxnet_gluon_mnist	2 months ago
<input type="checkbox"/>	mxnet_gluon_sentiment	2 months ago
<input type="checkbox"/>	mxnet_mnist	2 months ago
<input type="checkbox"/>	tensorflow_abalone_age_predictor_using_keras	2 months ago
<input type="checkbox"/>	tensorflow_abalone_age_predictor_using_layers	2 months ago
<input type="checkbox"/>	tensorflow_distributed_mnist	2 months ago
<input type="checkbox"/>	tensorflow_iris_dnn_classifier_using_estimators	2 months ago
<input type="checkbox"/>	tensorflow_resnet_cifar10_with_tensorboard	2 months ago
<input type="checkbox"/>	README.md	2 months ago

## End-to-End Example #1

1. [Introduction](#)
2. [Prerequisites and Preprocessing](#)
  - A. [Permissions and environment variables](#)
  - B. [Data ingestion](#)
  - C. [Data inspection](#)
  - D. [Data conversion](#)
3. [Training the K-Means model](#)
4. [Set up hosting for the model](#)
5. [Validate the model for use](#)

### Introduction

Welcome to our first end-to-end example! Today, we're working through a classification problem, specifically of images of handwritten digits, from zero to nine. Let's imagine that this dataset doesn't have labels, so we don't know for sure what the true answer is. In later examples, we'll show the value of "ground truth", as it's commonly known.



Not Trusted

conda\_python3

## Prerequisites and Preprocessing

### Permissions and environment variables

Here we set up the linkage and authentication to AWS services. There are two parts to this:

1. The role(s) used to give learning and hosting access to your data. Here we extract the role you created earlier for accessing your notebook. See the documentation if you want to specify a different role
2. The S3 bucket name and locations that you want to use for training and model data.

```
In [ ]: from sagemaker import get_execution_role  
        role = get_execution_role()  
        bucket='<bucket-name>'
```

### Data ingestion

Next, we read the dataset from the existing repository into memory, for preprocessing prior to training. In this case we'll use the MNIST dataset, which contains 70K 28 x 28 pixel images of handwritten digits. For more details, please see [here](#).

This processing could be done *in situ* by Amazon Athena, Apache Spark in Amazon EMR, Amazon Redshift, etc., assuming the dataset is present in the appropriate location. Then, the next step would be to transfer the data to S3 for use in training. For small datasets, such as this one, reading into memory isn't



labeled data (in the way we framed the problem), this is a good fit.

To get started, we need to set up the environment with a few prerequisite steps, for permissions, configurations, and so on.

## Prerequisites and Preprocessing

### Permissions and environment variables

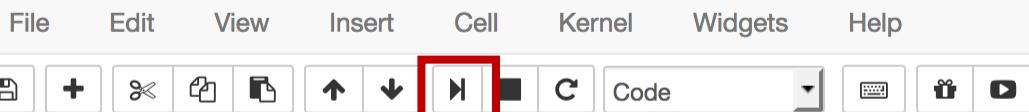
Here we set up the linkage and authentication to AWS services. There are two parts to this:

1. The role(s) used to give learning and hosting access to your data. Here we extract the role you created earlier for accessing your notebook. See the documentation if you want to specify a different role
2. The S3 bucket name and locations that you want to use for training and model data.

```
In [ ]: from sagemaker import get_execution_role  
  
role = get_execution_role()  
bucket='djenny-sagemaker-us-east-2'
```

### Data ingestion

Next, we read the dataset from the existing repository into memory, for preprocessing prior to training. In this case we'll use the MNIST dataset, which contains



Not Trusted

conda\_python3

labeled data (in the way we trained the problem), this is a good fit.

To get started, we need to set up the environment with a few prerequisite steps, for permissions, configurations, and so on.

## Prerequisites and Preprocessing

### Permissions and environment variables

Here we set up the linkage and authentication to AWS services. There are two parts to this:

1. The role(s) used to give learning and hosting access to your data. Here we extract the role you created earlier for accessing your notebook. See the documentation if you want to specify a different role
2. The S3 bucket name and locations that you want to use for training and model data.

```
In [ ]: from sagemaker import get_execution_role  
  
role = get_execution_role()  
bucket='djenny-sagemaker-us-east-2'
```

### Data ingestion

Next, we read the dataset from the existing repository into memory, for preprocessing prior to training. In this case we'll use the MNIST dataset, which contains 70K 28x28 pixel images of handwritten digits. For more details, see [here](#).



## Training the K-Means model

Once we have the data preprocessed and available in the correct format for training, the next step is to actually train the model using the data. Since this data is relatively small, it isn't meant to show off the performance of the k-means training algorithm. But Amazon SageMaker's k-means has been tested on, and scales well with, multi-terabyte datasets.

After setting training parameters, we kick off training, and poll for status until training is completed, which in this example, takes between 7 and 11 minutes.

```
In [ ]: from sagemaker import KMeans

data_location = 's3://{}{}/kmeans_highlevel_example/data'.format(bucket)
output_location = 's3://{}{}/kmeans_example/output'.format(bucket)

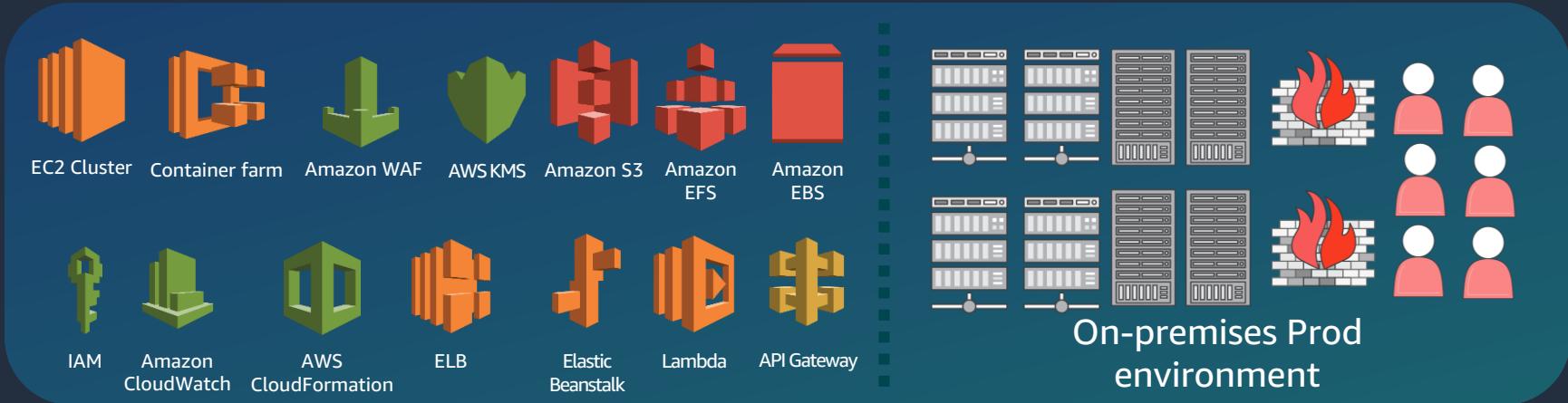
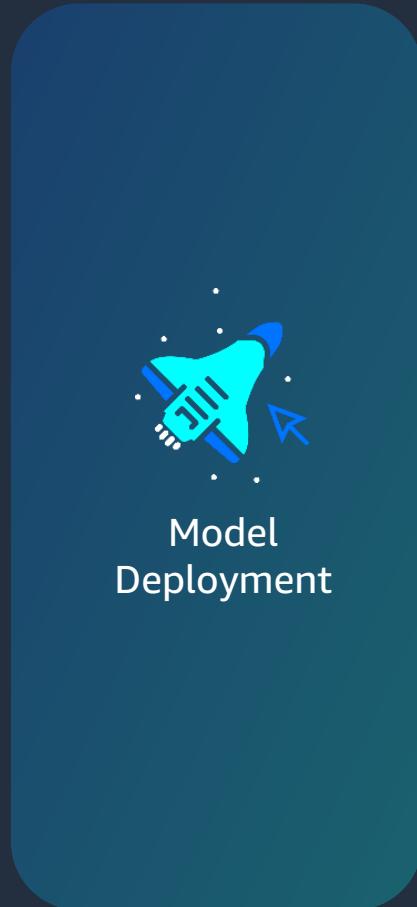
print('training data will be uploaded to: {}'.format(data_location))
print('training artifacts will be uploaded to: {}'.format(output_location))
```

```
kmeans = KMeans(role=role,
                  train_instance_count=2,
                  train_instance_type='ml.c4.8xlarge',
                  output_path=output_location,
                  k=10,
                  data_location=data_location)
```

```
In [ ]: %%time

kmeans.fit(kmeans.record_set(train_set[0]))
```

# Deploy Models - The Painful Way

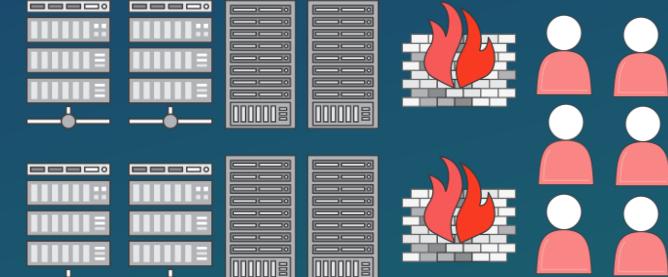


Deploy

# Deploy Models - The Painful Way



Model Deployment



On-premises Prod environment



One-click Deploy and Scaling

Deploy



Not Trusted

conda\_python3

## Set up hosting for the model

Now, we can deploy the model we just trained behind a real-time hosted endpoint. This next step can take, on average, 7 to 11 minutes to complete.

In [ ]:

```
%%time  
kmeans_predictor = kmeans.deploy(initial_instance_count=1,  
                                   instance_type='ml.m4.xlarge')
```

## Validate the model for use

Finally, we'll validate the model for use. Let's generate a classification for a single observation from the trained model using the endpoint we just created.

In [ ]:

```
result = kmeans_predictor.predict(train_set[0][30:31])  
print(result)
```

OK, a single prediction works.

Let's do a whole batch and see how well the clustering works.

In [ ]:

```
%%time
```



jupyter kmeans\_mnist Last Checkpoint: a few seconds ago (autosaved)



File Edit View Insert Cell Kernel Widgets Help

Trusted

conda\_python3



```
plt.show()
```

Cluster 0:



Cluster 1:



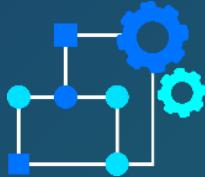
### The bottom line

K-Means clustering is not the best algorithm for image analysis problems, but we do see pretty reasonable clusters being built.

# Build Using SageMaker



Pre-built  
notebooks for  
common  
problems



Built-in, High  
performance  
Algorithms

---

**BUILD**

# Train Using SageMaker



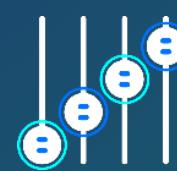
Pre-built  
notebooks for  
common  
problems



Built-in, High  
performance  
Algorithms



One-click  
Training



Hyperparameter  
Optimisation

---

BUILD

---

TRAIN

# Deploy Using SageMaker



Pre-built  
notebooks for  
common  
problems



Built-in, High  
performance  
Algorithms



One-click  
Training



Hyperparameter  
Optimisation



One-click  
Deployment



Fully managed  
hosting with  
Auto-scaling

**BUILD**

**TRAIN**

**DEPLOY**

# Improving Results

## More Data



1<sup>0</sup>0<sup>1</sup>0<sub>1</sub>  
1<sup>0</sup>0<sup>1</sup>0  
0<sup>1</sup>0<sup>1</sup>1  
1<sup>0</sup>1<sup>1</sup>0<sub>1</sub>

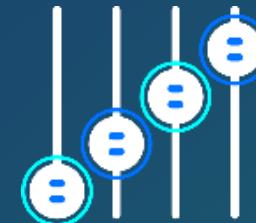
# Improving Results

More Data



1<sup>0</sup> 0<sup>1</sup> 0<sup>1</sup>  
0<sup>1</sup> 0<sup>1</sup> 0<sup>1</sup>  
0<sup>1</sup> 1<sup>0</sup> 1<sup>0</sup>  
1<sup>0</sup> 1<sup>1</sup> 0<sup>1</sup>

HPO



Hyperparameter  
Optimisation

# Improving Results

## More Data



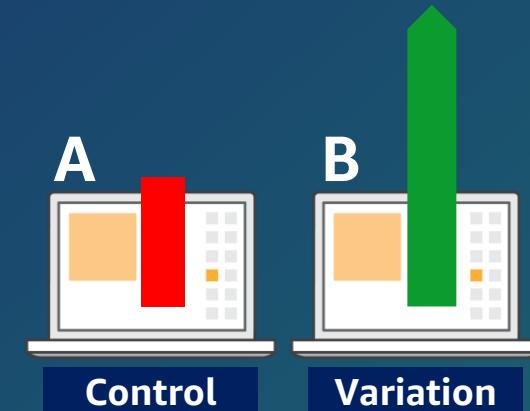
1<sup>0</sup>0<sup>1</sup>0<sup>1</sup>0<sup>1</sup>  
0<sup>1</sup>0<sup>1</sup>0<sup>1</sup>0<sup>1</sup>  
1<sup>0</sup>1<sup>1</sup>0<sup>1</sup>

## HPO



Hyperparameter  
Optimisation

## A/B Testing





# Learn from AWS experts. Advance your skills and knowledge. Build your future in the AWS Cloud.



**Digital Training**  
Free, self-paced online courses built by AWS experts



**Classroom Training**  
Classes taught by accredited AWS instructors



**AWS Certification**  
Exams to validate expertise with an industry-recognized credential

**Ready to begin building your cloud skills?**  
Get started at: <https://www.aws.training/>

# With deep expertise on AWS, APN Partners can help your organization at any stage of your Cloud Adoption Journey.



## AWS Managed Service Providers

APN Consulting Partners who are skilled at cloud infrastructure and application migration, and offer proactive management of their customer's environment.



## AWS Competency Partners

APN Partners who have demonstrated technical proficiency and proven customer success in specialized solution areas.



## AWS Marketplace

A digital catalog with thousands of software listings from independent software vendors that make it easy to find, test, buy, and deploy software that runs on AWS.



## AWS Service Delivery Partners

APN Partners with a track record of delivering specific AWS services to customers.

**Ready to get started with an APN Partner?**

**Find a partner:** <https://aws.amazon.com/partners/find/>

**Learn more at the AWS Partner Network Booth**

# Thank You for Attending AWS Innovate

We hope you found it interesting! A kind reminder to **complete the survey.**

Let us know what you thought of today's event and how we can improve the event experience for you in the future.

-  [aws-apac-marketing@amazon.com](mailto:aws-apac-marketing@amazon.com)
-  [twitter.com/AWSCloud](https://twitter.com/AWSCloud)
-  [facebook.com/AmazonWebServices](https://facebook.com/AmazonWebServices)
-  [youtube.com/user/AmazonWebServices](https://youtube.com/user/AmazonWebServices)
-  [slideshare.net/AmazonWebServices](https://slideshare.net/AmazonWebServices)
-  [twitch.tv/aws](https://twitch.tv/aws)