Agenda

09:00 Welcome and Introductions
09:15 Data Science Lifecycle
10:30 Lab 1: Understanding and preparing data with S3, Glue and Athena
11:15 Lunch
12:00 Model training, testing and deploying with Sagemaker
12:45 Lab 2: Train, test and deploy your first model with Sagemaker
13:45 Break
14:00 Continuous Delivery of ML models
14:30 Lab 3: Continuous Delivery of ML models to Amazon SageMaker
15:15 Bring your own model
15:45 Wrap Up

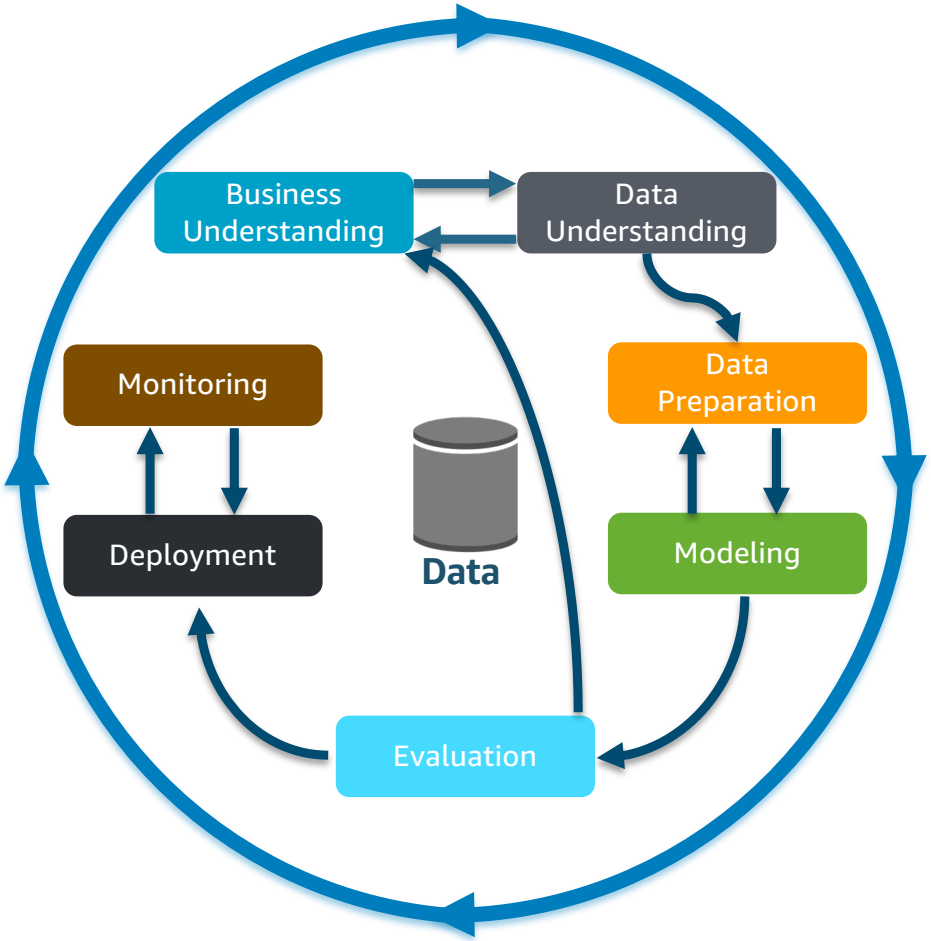
aws

Model training, testing and deploying with Amazon SageMaker

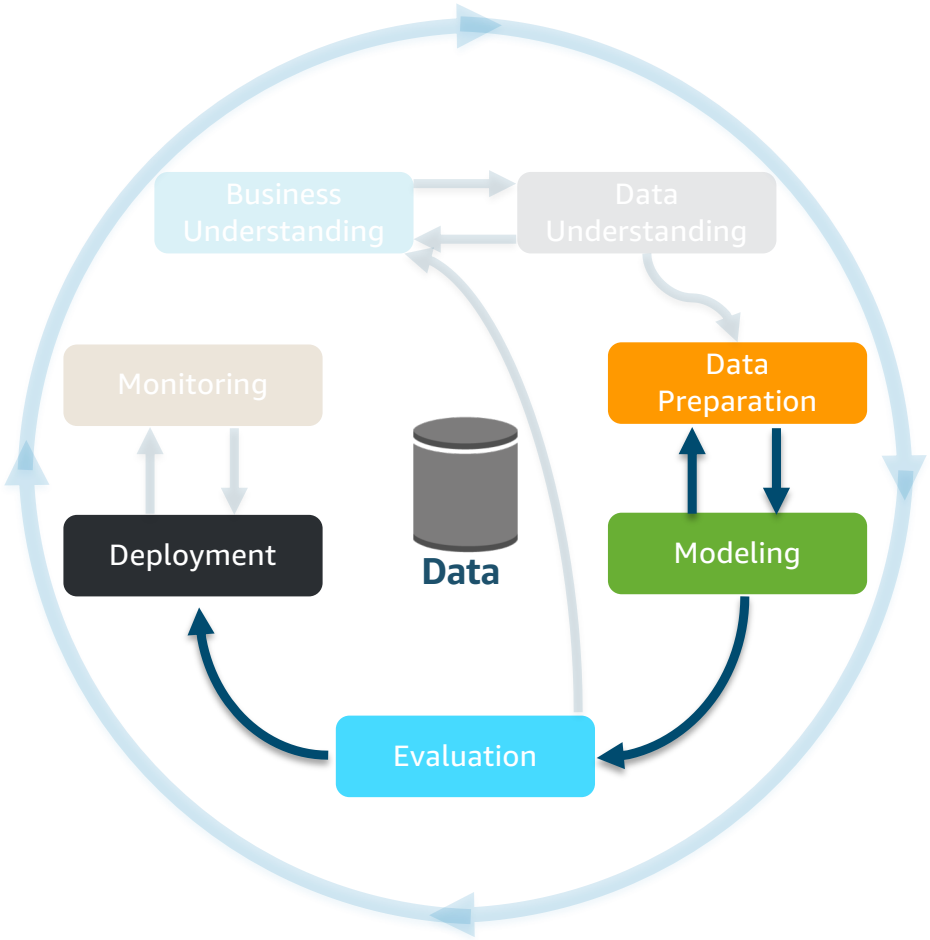
Machine Learning Immersion Day

Module 2

Data Science Lifecycle

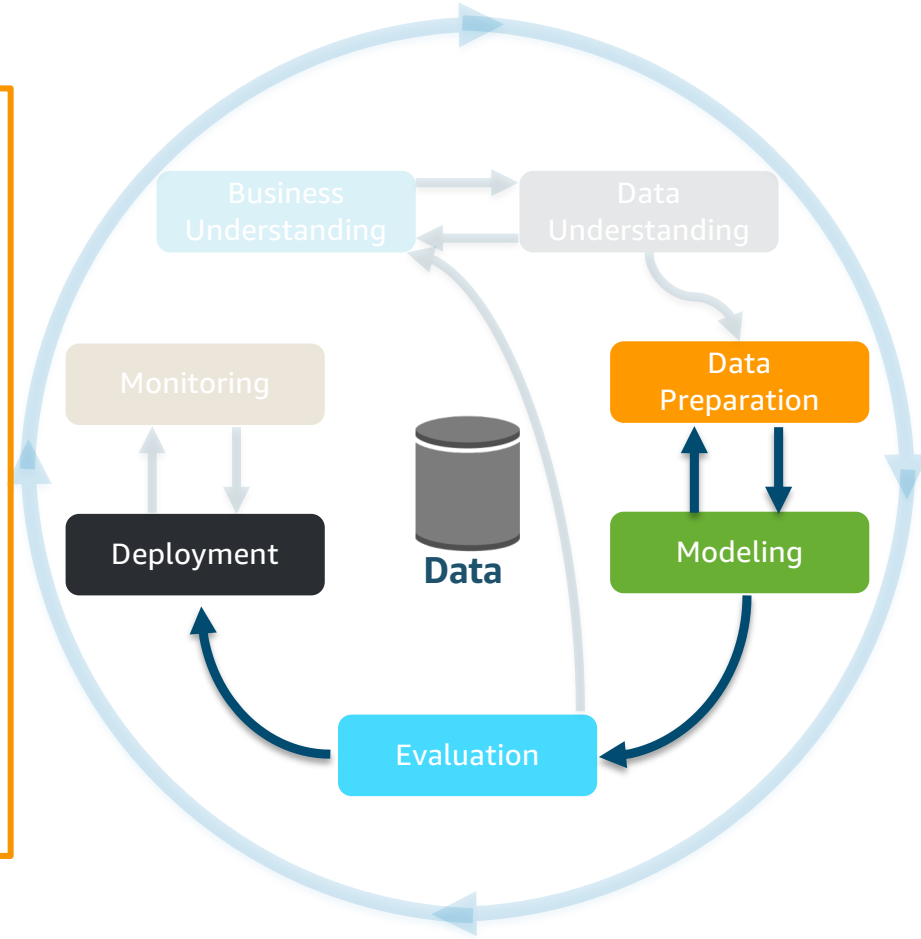


Data Science Lifecycle



ML Heavy Lifting

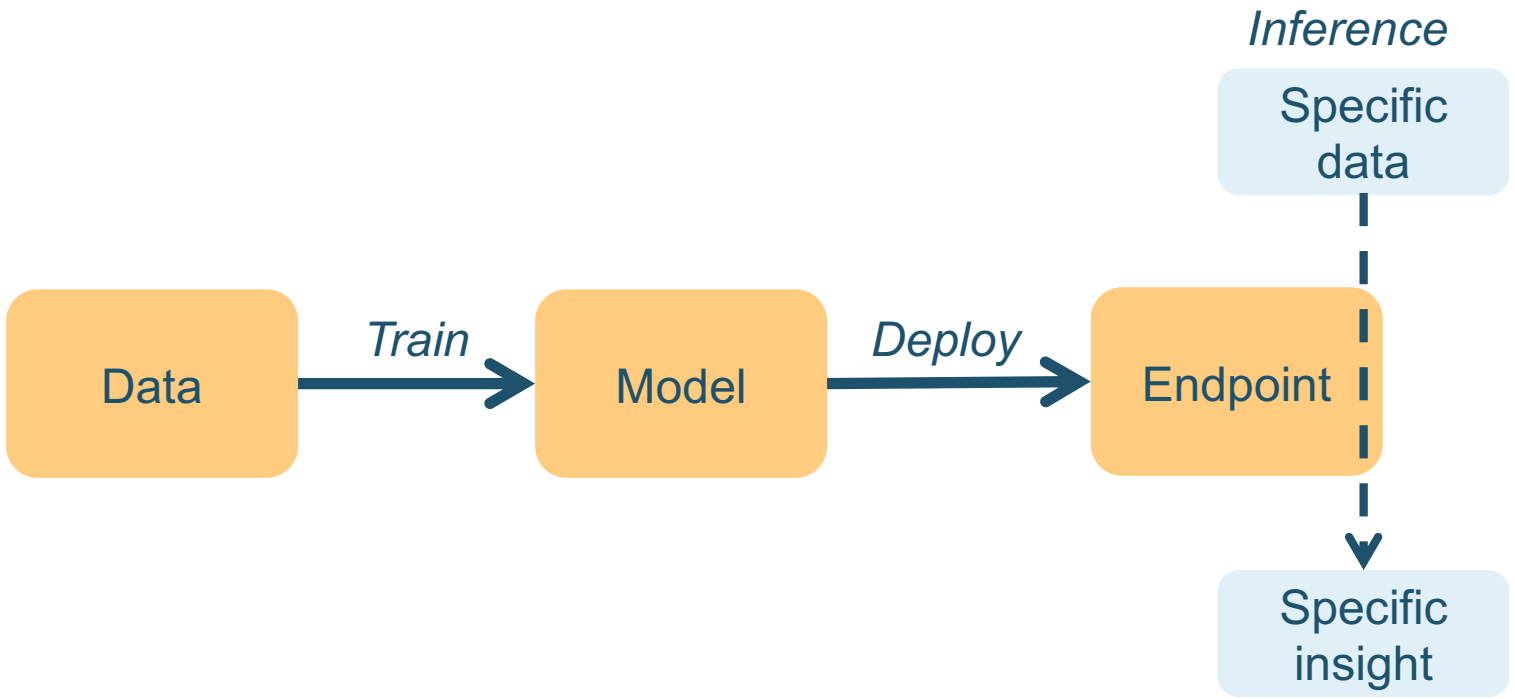
- Setup and manage Model Inference Clusters
- Manage and Scale Model Inference APIs
- Monitor and Debug Model Predictions
- Models versioning and performance tracking
- Automate New Model version promotion to production (A/B testing)



- Setup and manage Notebook Environments
- Setup and manage Training Clusters
- Write Data Connectors
- Scale ML algorithms to large datasets
- Distribute ML training algorithm to multiple machines
- Secure Model artifacts

aws

ML terminology

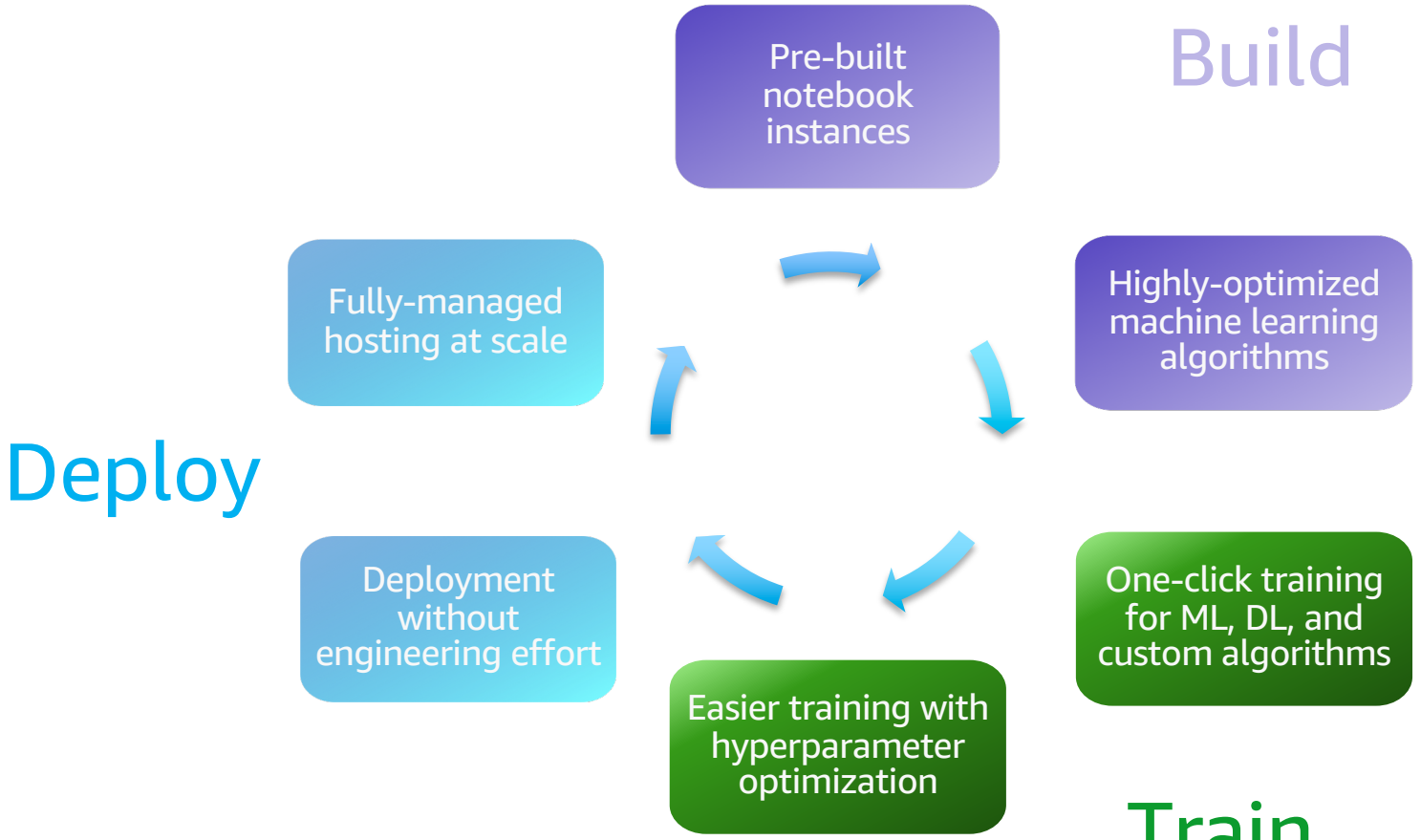


Amazon SageMaker

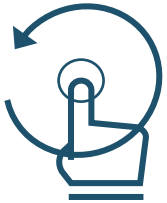
A managed service
that provides **the quickest and easiest way** for
data scientists and developers to get
ML models from idea to production.

aws

Amazon SageMaker

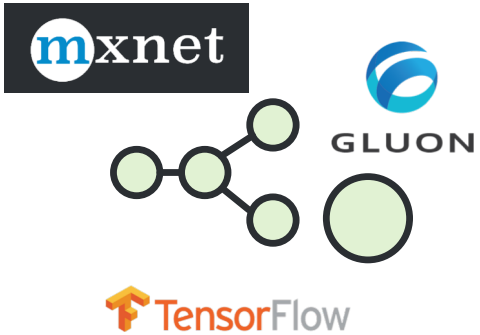


Amazon SageMaker

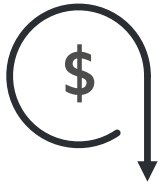


End-to-End Machine Learning Platform

Zero setup



Flexible Model Training



Pay by the second

aws

Amazon SageMaker components - UX

UX | **Training** | **Hosting**

Amazon's fast, scalable algorithms

Idiomatic, distributed TensorFlow & MXNet

Bring your own algorithm

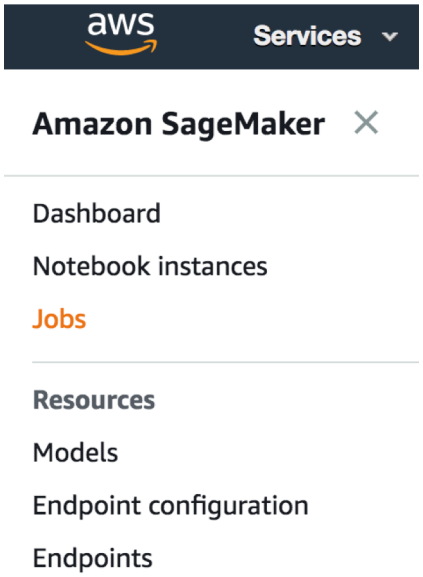
Hyperparameter optimization

aws

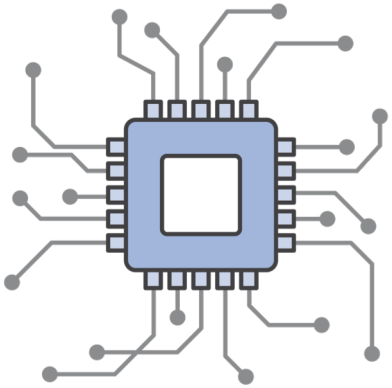
UX

Use SageMaker's hosted Notebook Instances…

… or Apache Spark through EMR and the SageMaker Spark SDK…



… or SageMaker's Console for a point and click experience…



… or your own device (EC2, laptop, etc.)

aws

Amazon SageMaker components - training

UX | **Training** | **Hosting**

Amazon's fast, scalable algorithms

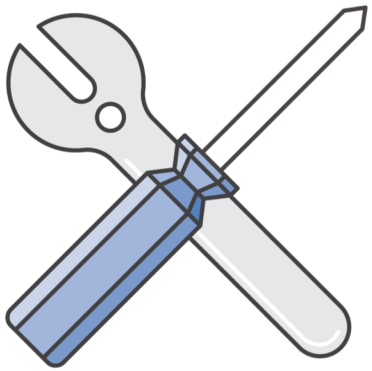
Idiomatic, distributed TensorFlow & MXNet

Bring your own algorithm

Hyperparameter optimization

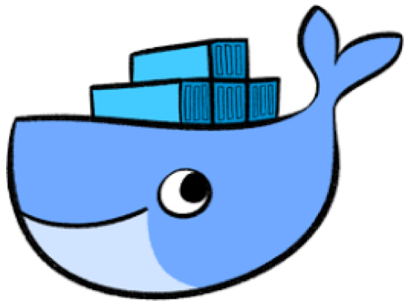
aws

Training



Zero setup

Streaming datasets + distributed compute



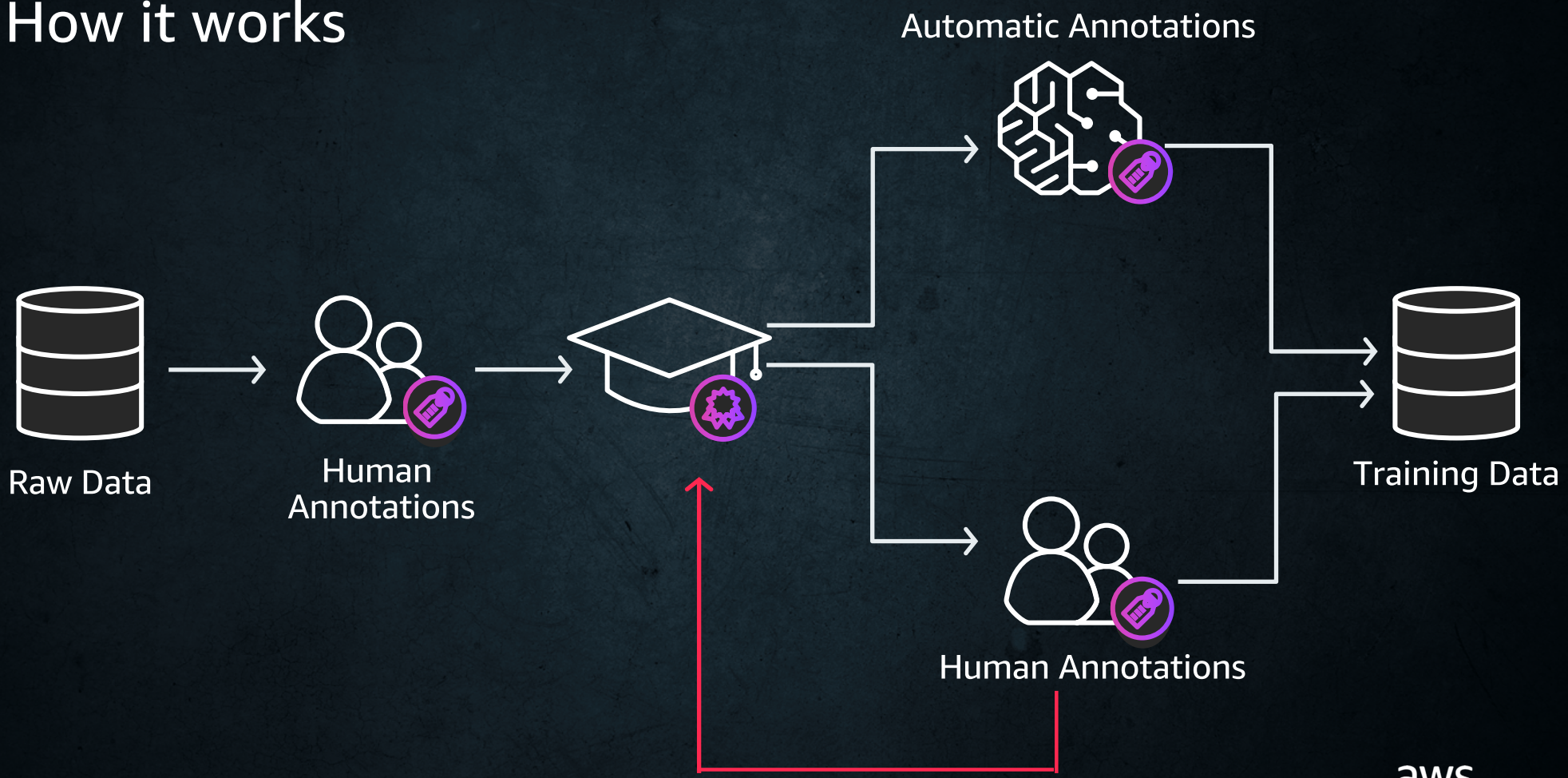
Docker / ECS

Deploy trained models locally or to SageMaker, Greengrass, DeepLens

aws

Amazon SageMaker Ground Truth
How it works

NEW

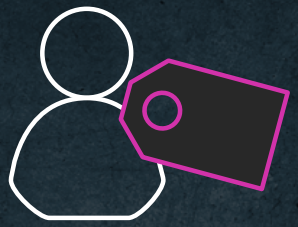


aws

NEW

Amazon SageMaker Ground Truth
Label machine learning training data easily and accurately

Quickly label training data



Easily integrate human labelers

Get accurate results

KEY FEATURES

- Automatic labeling via machine learning
- Ready-made and custom workflows
- Private and public human workforce
- Label management

Tyson

CHANGE HEALTHCARE

Automagi

gumgum

Amazon SageMaker components - hosting

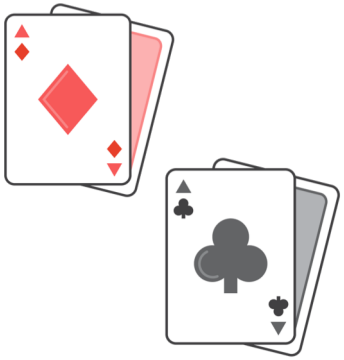
UX	Training	Hosting
	Amazon's fast, scalable algorithms	
	Idiomatic, distributed TensorFlow & MXNet	
	Bring your own algorithm	
Hyperparameter optimization		

aws

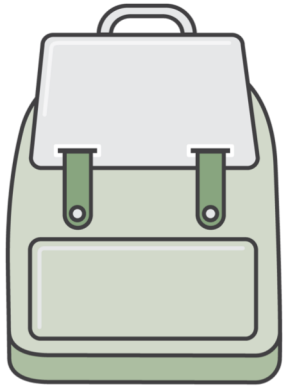
Hosting

One step deployment

Low latency, high throughput, and high reliability



A/B testing



Use your own model

aws

Amazon SageMaker components – Built-in Algorithms

UX | **Training** | **Hosting**

Amazon's fast, scalable algorithms

Idiomatic, distributed TensorFlow & MXNet

Bring your own algorithm

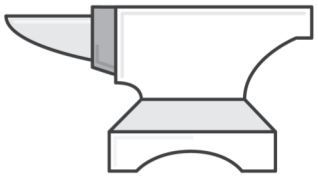
Hyperparameter optimization

aws

Amazon SageMaker

Streaming datasets, for cheaper training

Train faster, in a single pass

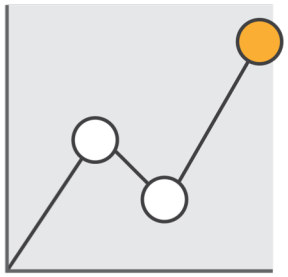


Greater reliability on extremely large datasets

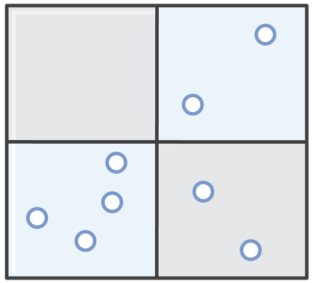
Choice of several ML algorithms

aws

Built-in algorithms



XGBoost, FM, Linear, and Forecasting for supervised learning



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

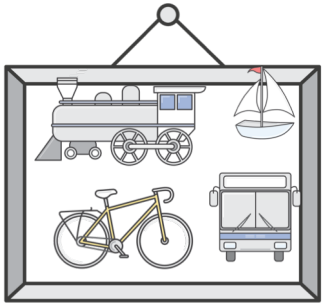
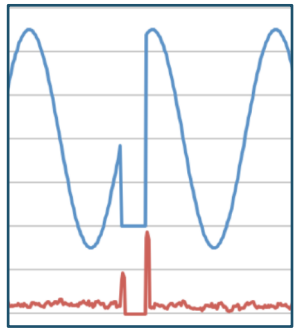


Image classification with convolutional neural networks

LDA and NTM for topic modeling, seq2seq for translation



Random Cut Forest for anomaly detection

aws

Amazon SageMaker components – Hyperparameter optimization

UX | **Training** | **Hosting**

Amazon's fast, scalable algorithms

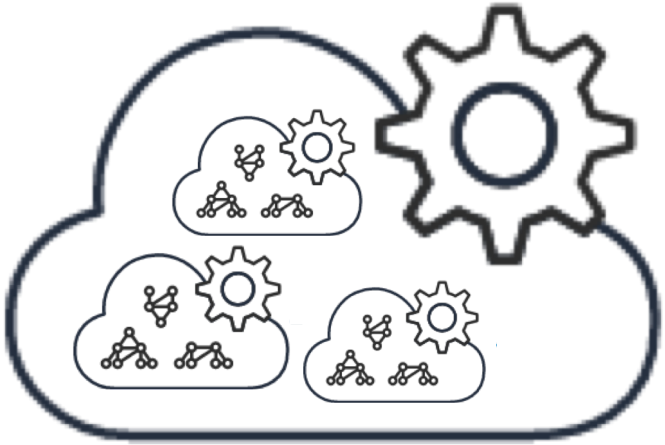
Idiomatic, distributed TensorFlow & MXNet

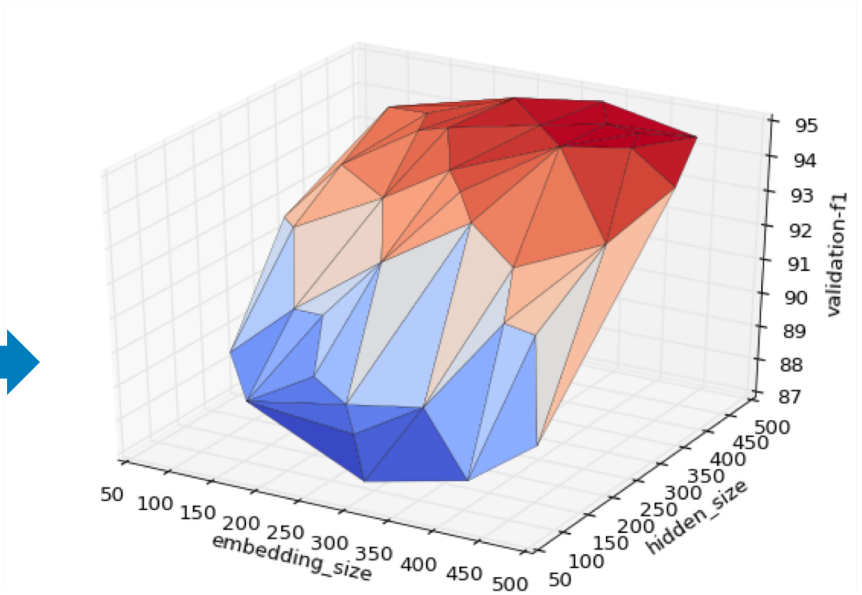
Bring your own algorithm

Hyperparameter optimization

aws

Hyperparameter Optimization



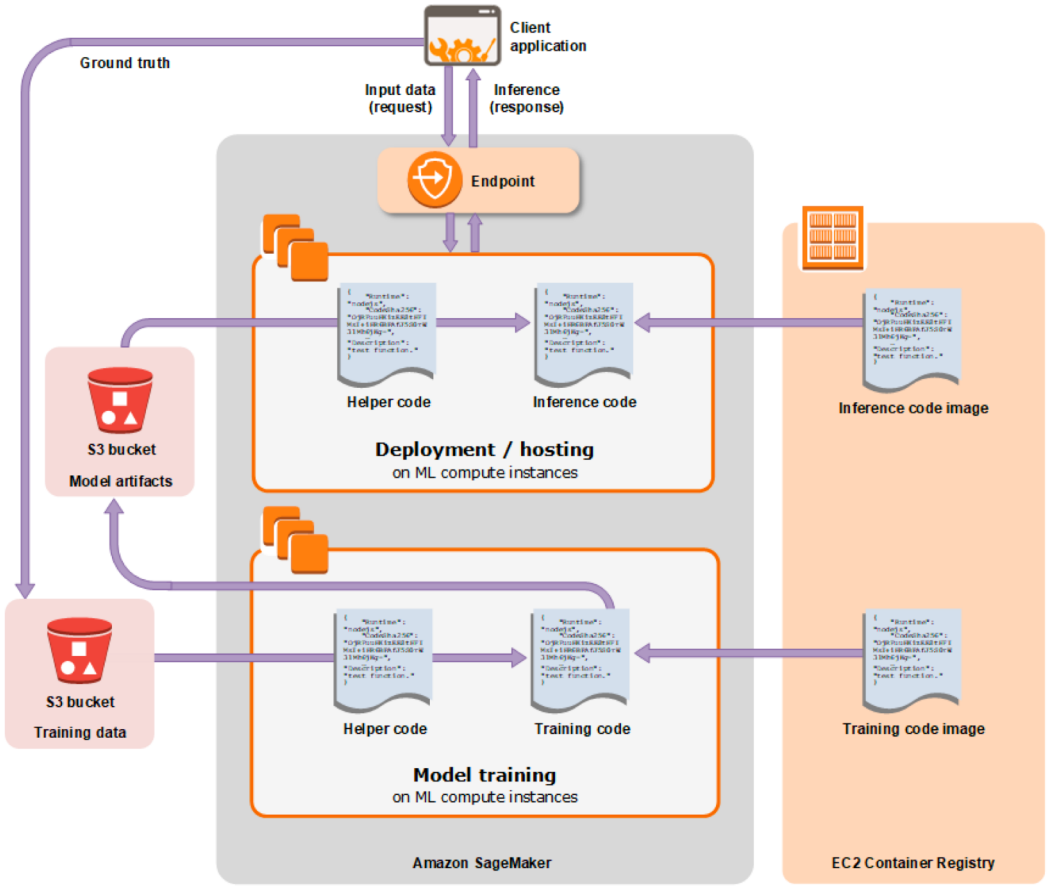


Run a large set of training jobs with varying hyperparameters...

... and search the hyperparameter space for improved accuracy.

aws

From training to inference



Data preparation

Recommender systems and Factorization Machines

- Enhance user experience using a recommender system based on ratings from previous users
- Often many users and many items but few recommendations
- FMs efficient on large sparse datasets
- Linear training time
- Factorize recommendation matrix

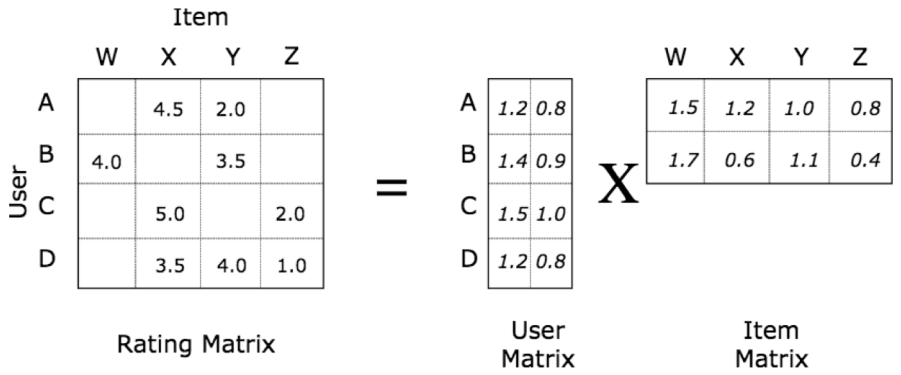


Image source: http://data-artisans.com

aws

Data preparation

1. Separate MovieLens data to training and test set
2. For each set, build a sparse matrix holding one-hot encoded data samples.
3. For each set, build a label vector holding ratings.
4. Write both sets to protobuf-encoded files.
5. Copy these files to an Amazon S3 bucket.

aws

One-hot encoding

- Mark 1 for one user and one movie per line in sparse feature matrix.
- Mark 1 if rating >=4, else 0 in label vector

User	Movie	Rating
2	2	4
1	3	5
3	1	4
1	1	2

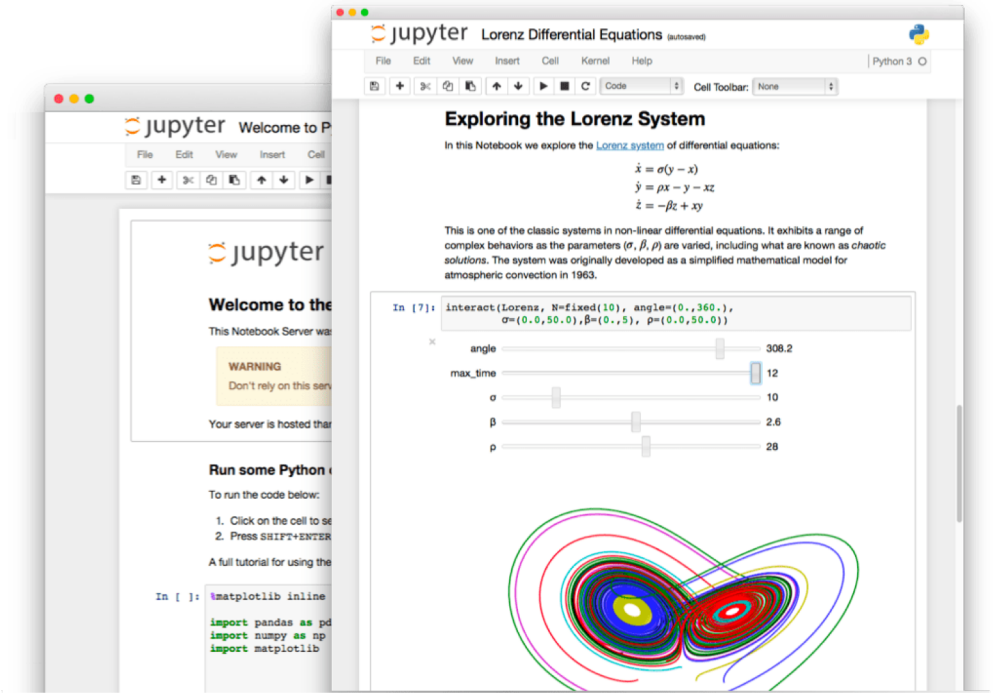
User 1	User 2	User 3	...	User M	Movie 1	Movie 2	Movie 3	...	Movie N
	1					1			
1							1		
		1			1				
1					1				

Label
1
1
1
0

aws

Jupyter notebooks

- Web-based interactive computational environment
- Collects code, text (Markdown), mathematics, plots
- Ordered input/output cells



aws

Questions?