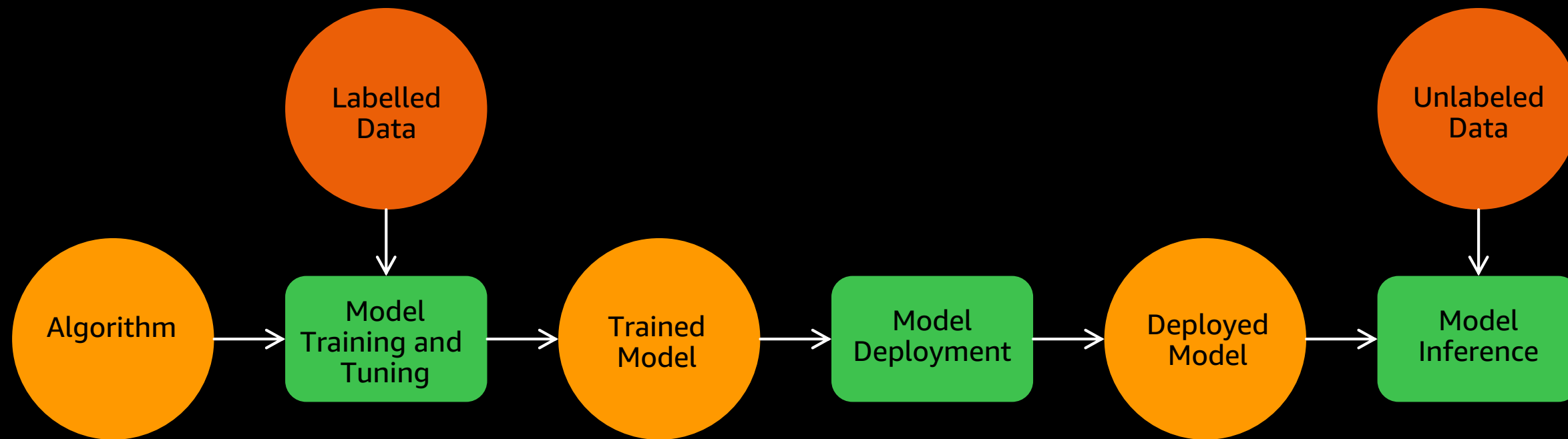


Jumpstart to prepare, build, train, and deploy ML models on AWS























Tapan Hoskeri
Solutions Architect
AISPL

A simple machine learning workflow



The AWS ML Stack

Broadest and most complete set of machine learning capabilities

AI SERVICES			HEALTH AI			INDUSTRIAL AI			ANOMALY DETECTION		CODE AND DEVOPS	
												
NEW			NEW	NEW	NEW	NEW	NEW	NEW	NEW			
Amazon HealthLake	Amazon Transcribe for Medical	Amazon Comprehend for Medical	AWS Panorama + Appliance	Amazon Monitron	Amazon Lookout for Equipment	Amazon Lookout for Vision	Amazon Lookout for Metrics	Amazon DevOps Guru	Amazon CodeGuru			
VISION	SPEECH		TEXT			SEARCH	CHATBOTS	PERSONALIZATION	FORECASTING	FRAUD	CONTACT CENTERS	
												
Amazon Rekognition	Amazon Polly	Amazon Transcribe <i>+Medical</i>	Amazon Comprehend <i>+Medical</i>	Amazon Translate	Amazon Textract	Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast	Amazon Fraud Detector	Contact Lens <i>For Amazon Connect</i>	

ML SERVICES

Amazon SageMaker

Label data

NEW Aggregate & prepare data	NEW Store & share features	Auto ML	Spark/R	NEW Detect bias	Visualize in notebooks	Pick algorithm	Train models	Tune parameters	NEW Debug & profile	Deploy in production	Manage & monitor	NEW CI/CD	Human review
--	--------------------------------------	---------	---------	---------------------------	------------------------	----------------	--------------	-----------------	-------------------------------	----------------------	------------------	---------------------	--------------

SAGEMAKER STUDIO IDE

NEW: SageMaker JumpStart

NEW: Model management for edge devices

FRAMEWORKS & INFRASTRUCTURE

TensorFlow	mxnet	GLUON	Keras	Deep Learning AMIs & Containers	GPUs & CPUs	Elastic Inference	Trainium	Inferentia	FPGA
PyTorch	scikit learn	HOROVOD	DeepGraphLibrary						



Amazon SageMaker overview

Amazon SageMaker

PREPARE

SageMaker Ground Truth

Label training data for machine learning

SageMaker Data Wrangler **NEW**

Aggregate and prepare data for machine learning

SageMaker Processing

Built-in Python, BYO R/Spark

SageMaker Feature Store **NEW**

Store, update, retrieve, and share features

SageMaker Clarify **NEW**

Detect bias and understand model predictions

BUILD

SageMaker Studio Notebooks

Jupyter notebooks with elastic compute and sharing

Built-in and Bring your-own Algorithms

Dozens of optimized algorithms or bring your own

Local Mode

Test and prototype on your local machine

SageMaker Autopilot

Automatically create machine learning models with full visibility

SageMaker JumpStart **NEW**

Pre-built solutions for common use cases

TRAIN & TUNE

Managed Training

Distributed infrastructure management

SageMaker Experiments

Capture, organize, and compare every step

Automatic Model Tuning

Hyperparameter optimization

Distributed Training **NEW**

Training for large datasets and models

SageMaker Debugger **NEW**

Debug and profile training runs

Managed Spot Training

Reduce training cost by 90%

DEPLOY & MANAGE

Managed Deployment

Fully managed, ultra low latency, high throughput

Kubernetes & Kubeflow Integration

Simplify Kubernetes-based machine learning

Multi-Model Endpoints

Reduce cost by hosting multiple models per instance

SageMaker Model Monitor

Maintain accuracy of deployed models

SageMaker Edge Manager **NEW**

Manage and monitor models on edge devices

SageMaker Pipelines **NEW**

Workflow orchestration and automation

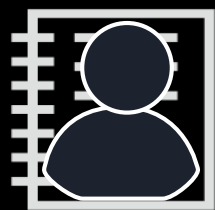
SageMaker Studio

Integrated development environment (IDE) for ML



Amazon SageMaker Studio

Fully integrated development environment (IDE) for machine learning



Collaboration at scale

Share notebooks without tracking code dependencies



Easy experiment management

Organize, track, and compare thousands of experiments



Automatic model generation

Get accurate models with full visibility & control without writing code



Higher quality ML models

Automatically debug errors, monitor models, & maintain high quality



Increased productivity

Code, build, train, deploy, & monitor in a unified visual interface



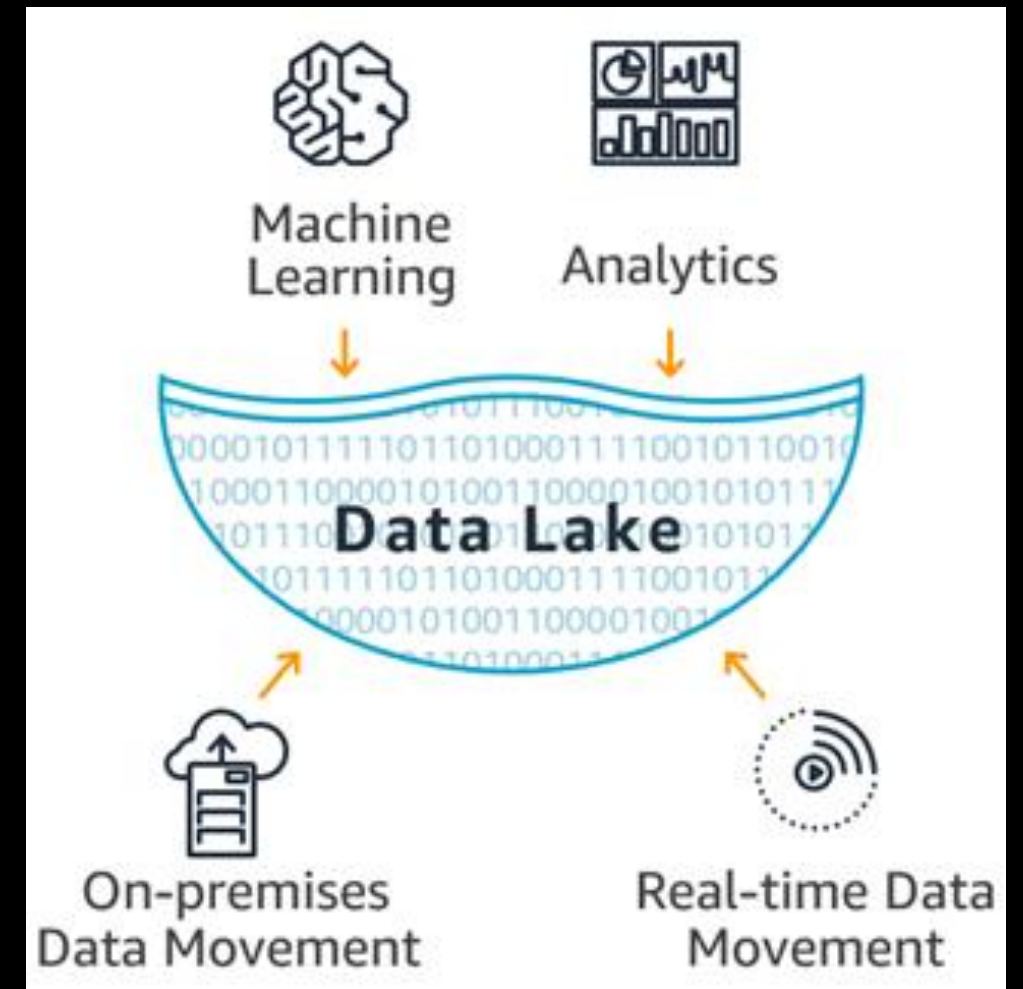
Data

What is a Data Lake?

Centralized repository that allows you to store all your structured and unstructured data at any scale.

Store your data as-is, without having to first structure the data

Run different types of analytics—from dashboards and visualizations to big data processing, real-time analytics, and Machine Learning

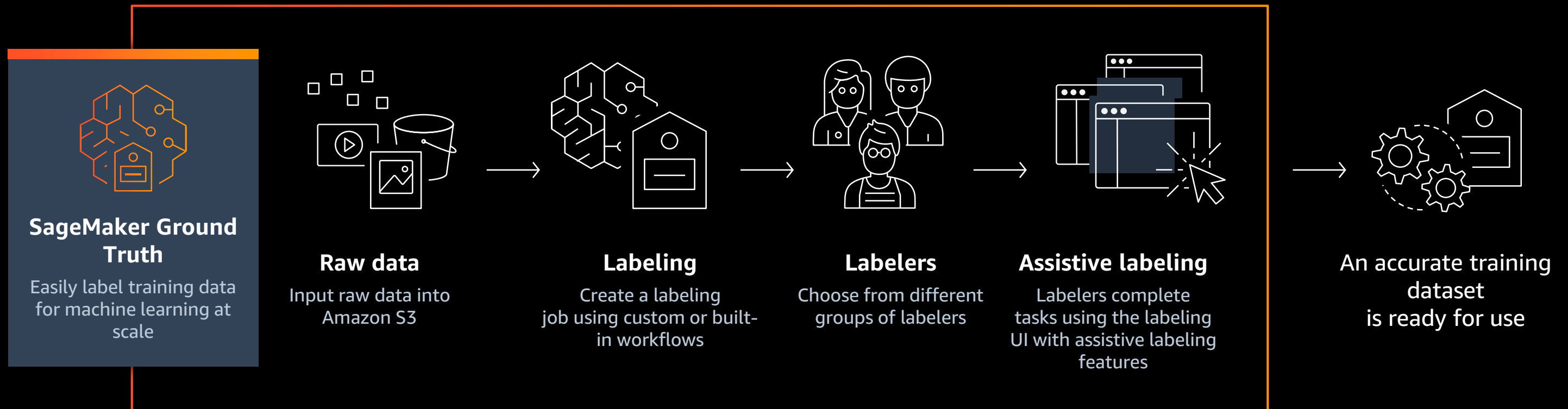


Amazon Simple Storage Service (Amazon S3)

- Object storage service object storage service that offers industry-leading scalability, data availability, security, and performance.
- Designed for 99.999999999% (11 9's) of durability.
- Scale storage resources up and down to meet fluctuating demands, without upfront investments or resource procurement cycles.
- Stores data for millions of applications for companies all around the world.



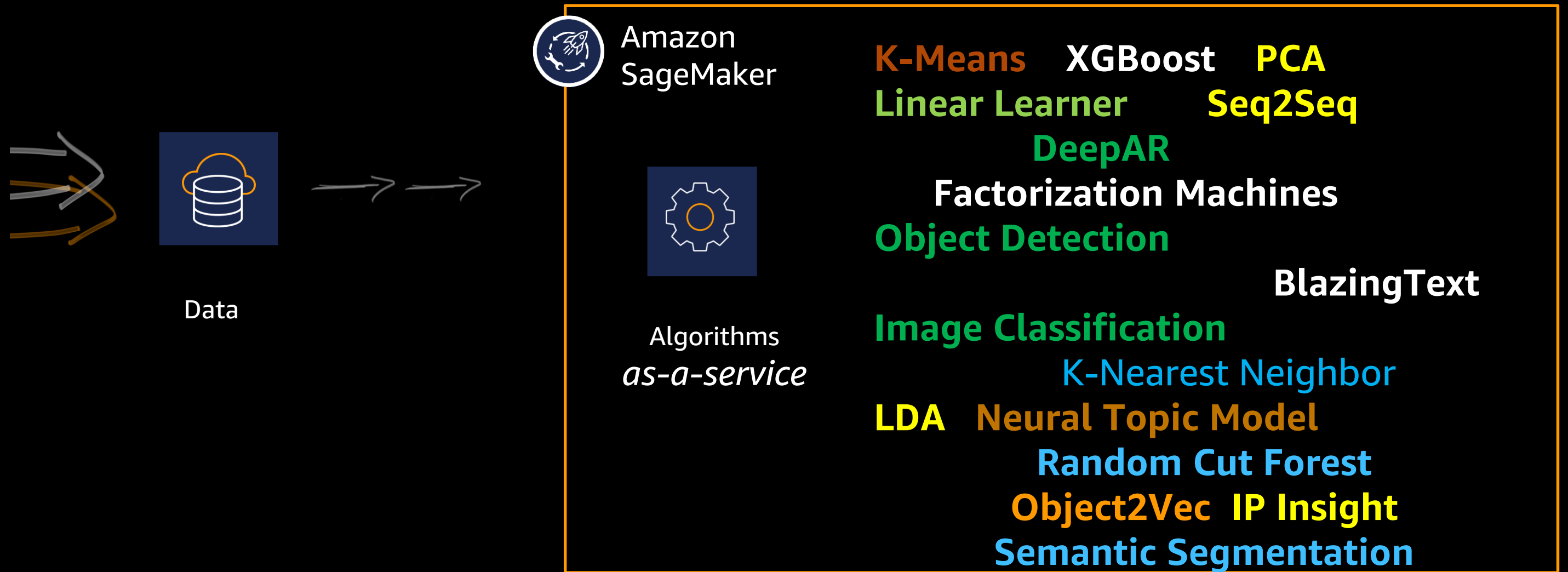
How Amazon SageMaker Ground Truth works



Algorithms

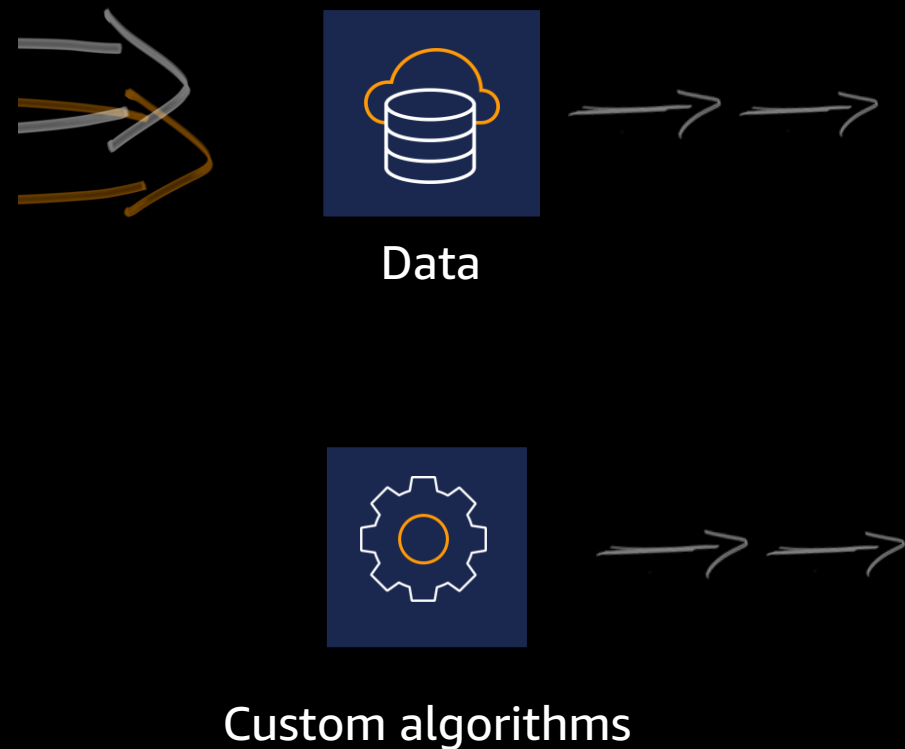
Amazon SageMaker – Built-in Algorithms

Rethinking algorithms design for large scale & streaming data



Amazon SageMaker – Bring-Your-Own-Algorithm

Build your own algorithms, SageMaker handles the rest



Amazon SageMaker Training Service Supported Containers



Amazon SageMaker – Bring-Your-Own-Container

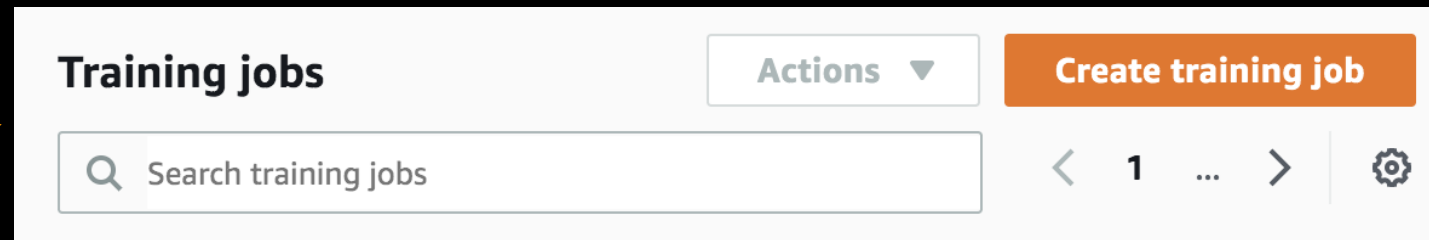
Bring your custom code and container, train at scale in SageMaker



Training your model

Launching Training Jobs from Jupyter

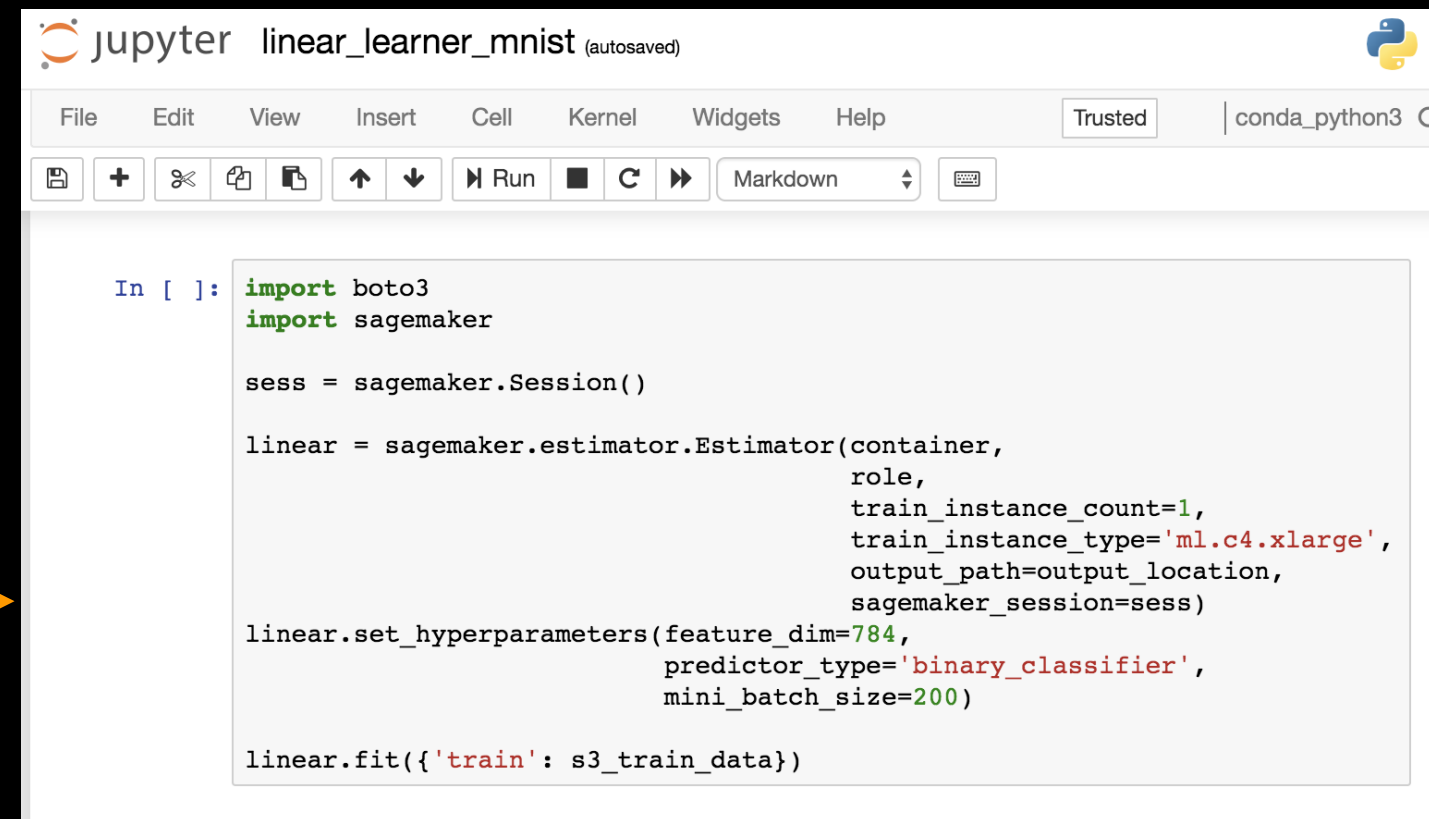
One click in Console



- OR -

Launch Training

Using API/SDK



The screenshot shows a Jupyter Notebook interface with the title 'linear_learner_mnist (autosaved)'. The code in the notebook is as follows:

```
In [ ]: import boto3
import sagemaker

sess = sagemaker.Session()

linear = sagemaker.estimator.Estimator(container,
                                       role,
                                       train_instance_count=1,
                                       train_instance_type='ml.c4.xlarge',
                                       output_path=output_location,
                                       sagemaker_session=sess)

linear.set_hyperparameters(feature_dim=784,
                           predictor_type='binary_classifier',
                           mini_batch_size=200)

linear.fit({'train': s3_train_data})
```



Compute instances for each workload

Well suited for notebooks

Well suited for AI/ML training workloads



Well suited for AI/ML simulation and inference workloads



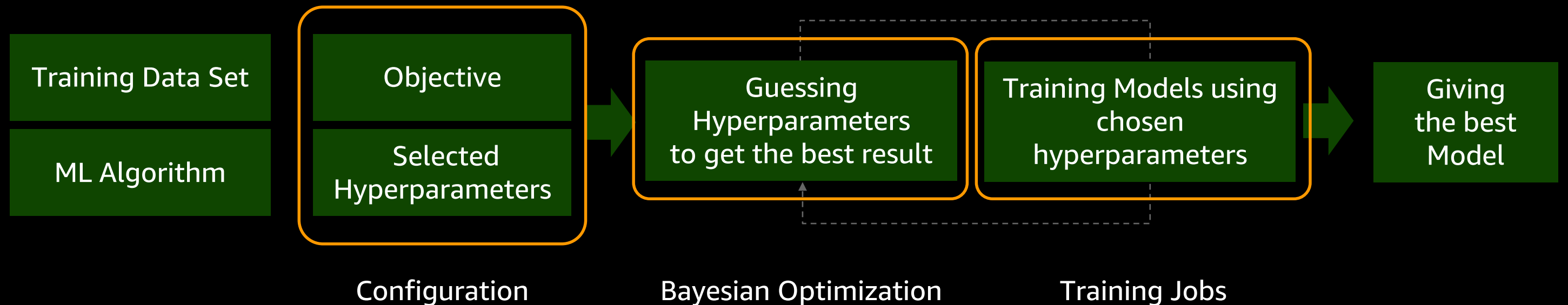
Tuning your model

Automatic Model Tuning

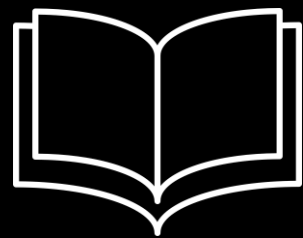
Hyperparameter tuning finds the best hyperparameter using Bayesian optimization

All ML Algorithms Supported

SageMaker Built-in Algorithms or custom algorithms



Amazon SageMaker Autopilot



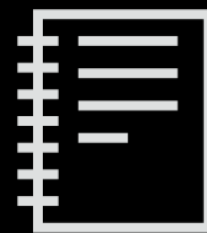
Quick to start

Provide your data in a tabular form & specify target prediction



Automatic model creation

Get ML models with feature engineering & model tuning automatically done



Visibility & control

Get notebooks for your models with source code

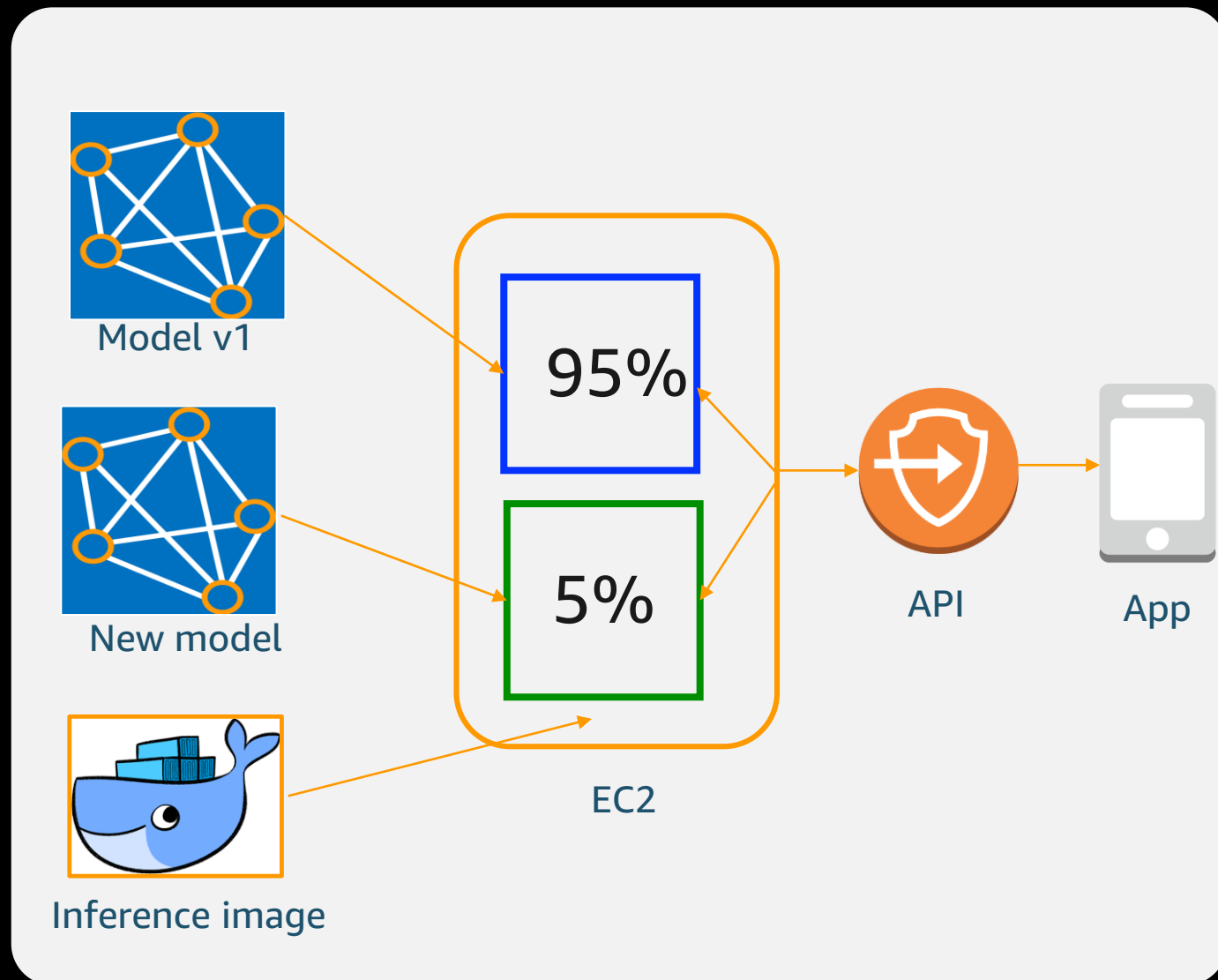


Recommendations & optimization

Get a leaderboard & continue to improve your model

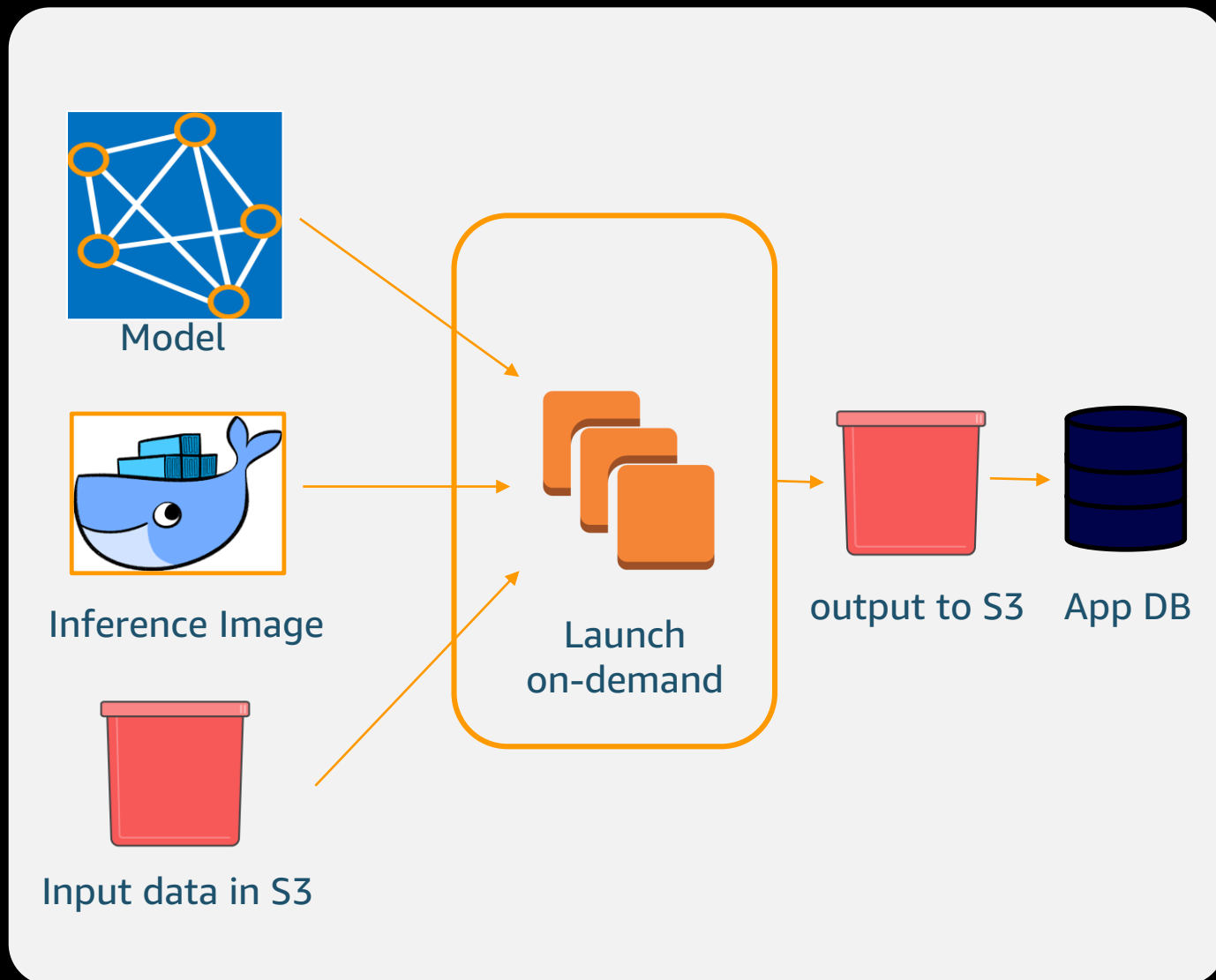
Deploying your model

Amazon SageMaker – Real time inference



- Quick deployment of Restful API endpoint
- Autoscaling across multiple AZs for variable loads
- A/B testing support
- VPC privatelink support
- Endpoint invocation and resource usage metrics
- No model re-engineering

Amazon SageMaker – Batch inference



- Use trained model to get inference on data in S3
- Launch fully managed transient inference resources
- Seamless integration with Amazon S3 for both input and output

Cost optimization

Managed spot training

Up to 90% lower compute costs

If a spot instance is interrupted, Amazon SageMaker automatically resumes the training job when spot instances become available

Managed spot training



Enable managed spot training - *optional*

Save compute costs for jobs that have flexibility in start and end times. Amazon SageMaker will use spare capacity only to run this job. [Learn more](#)

Maximum wait time before job terminates *optional stopping condition*

At the end of this duration you will receive the complete or partial results of you managed spot training job.



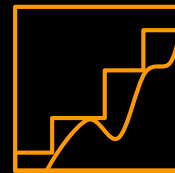
Amazon Elastic Inference

Reduce deep learning inference costs up to 75%



Lower
inference costs

Integrated with
Amazon EC2 and
Amazon SageMaker



Match capacity
to demand

Support for TensorFlow,
Apache MXNet (Incubating)
—PyTorch coming soon



Available between
1 to 32 TFLOPS
per accelerator

Single and mixed-
precision operations





Demo

Visit the AI and Machine Learning Resource Hub for more resources

Dive deeper with these resources, get inspired and learn how you can use machine learning to accelerate business outcomes.

- The machine learning journey e-book
- Machine learning enterprise guide
- 7 leading machine learning use cases e-book
- A strategic playbook for data, analytics, and machine learning
- Accelerating ML innovation through security e-book
- ... and more!

[Visit resource hub »](#)



<https://tinyurl.com/aiml-aws>



AWS Machine Learning (ML) Training and Certification

Learn like an Amazonian, based on the curriculum we've used to train our own developers and data scientists



AWS is how you build machine learning skills

Courses built on the curriculum leveraged by Amazon's own teams. Learn from the experts at AWS.



Flexibility to learn your way

Learn online with 65+ on-demand digital courses or live with virtual instructor-led training, plus hands-on labs and opportunities for practical application.



Validate your expertise

Demonstrate expertise in building, training, tuning, and deploying machine learning models with an industry-recognized credential.

aws.training/machinelearning



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



twitter.com/AWSCloud



facebook.com/AmazonWebServices



youtube.com/user/AmazonWebServices



slideshare.net/AmazonWebServices



twitch.tv/aws



Thank you!