Amazon SageMaker

Amazon SageMaker is a fully managed service for preparing, building, training, and deploying high-quality machine learning models.
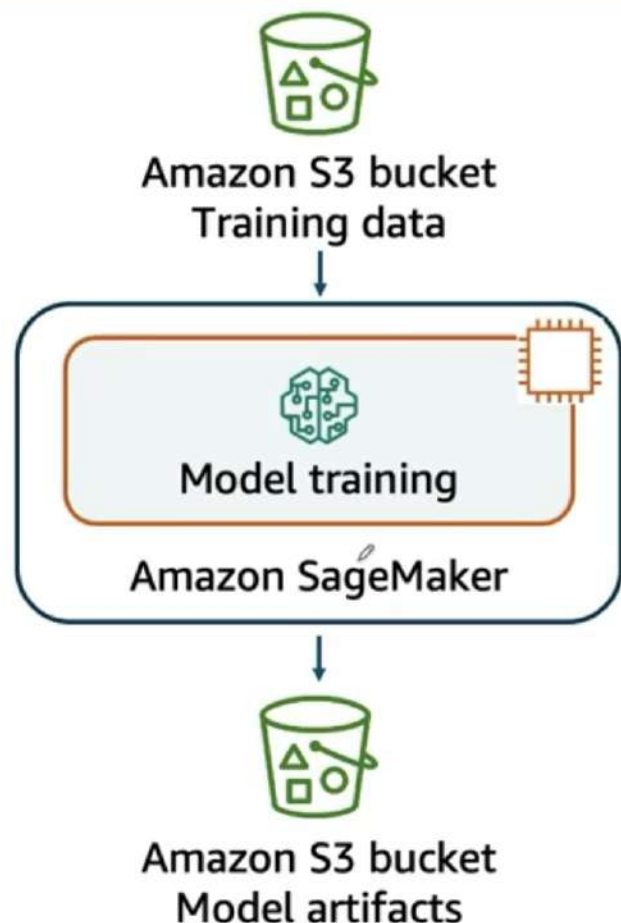
Amazon SageMaker

Notebook instances

An *Amazon SageMaker notebook instance* is a machine learning instance, where you can compile and share code, documentation, and visualization models.

With notebook instances in Amazon SageMaker, you can do the following:

- Launch managed instances for Jupyter Notebooks.

- Prepare and process data.

- Write code to train models.

- Deploy and validate models.

From SageMaker, we can launch managed notebook instances.

Training jobs

Amazon S3 bucket
Training data

Model training

Amazon SageMaker

Amazon S3 bucket
Model artifacts

Algorithm options
- Amazon SageMaker algorithms: XGBoost, Object2Vec, K-Means, and more
- You can supply your own algorithm
- AWS Marketplace algorithms

Resource configurations
- Standard
- Compute optimized
- Accelerated instance types

Hyperparameters
- Fine control over algorithm settings

Input data channels
- Training and validation data

aws

AWS Marketplace

Buy and sell Amazon SageMaker *algorithms* for training and inference.

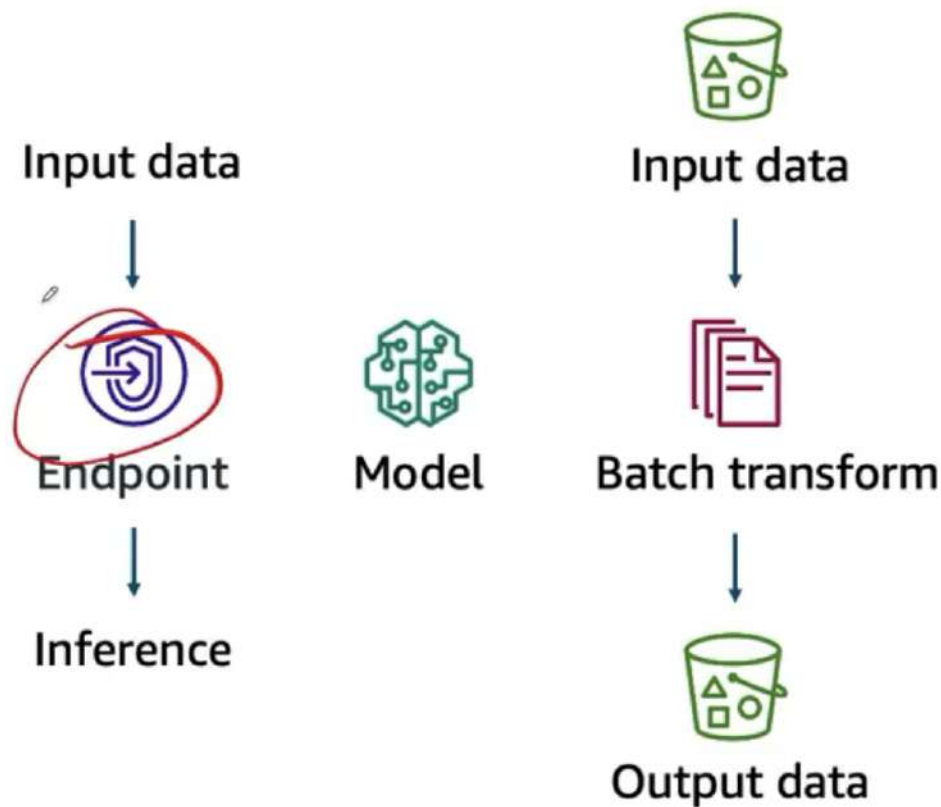Buy and sell Amazon SageMaker *model packages* for inference.

On the previous slide, we talked about the algorithms

Inference

Input data

↓

Endpoint

↓

Inference

Model

Input data

↓

Batch transform

↓

Output data

To generate inferences from a model, use:

- Real-time inference
- Serverless inference
- Asynchronous inference
- Batch transform

aws   Services   Q Search for services, features, blogs, docs, and more   [Option+S]   Oregon ▼   machine-learning ▼

Resource Groups & Tag Editor   S3   CloudFormation

Amazon SageMaker   ×

Getting started

Control panel
  Studio
  Studio Lab ☒
  Canvas
  RStudio

SageMaker dashboard
  Images
  Lifecycle configurations
  Search
▶ Ground Truth
▶ Notebook
▶ Processing
▶ Training
▼ Inference
  Compilation jobs
  Marketplace model packages
  Models
  Endpoint configurations

MACHINE LEARNING

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

What is Amazon SageMaker?

Amazon SageMaker provides machine learning (ML) capabilities for data scientists and developers to prepare, build, train, and deploy high-quality ML models efficiently.

New user onboarding guide   NEW

Get started with Amazon SageMaker by completing the quick start onboarding guide.

Get started with SageMaker

New to SageMaker?

Get started with Amazon SageMaker by completing the quick start guide.

Get Started

Documentation

Getting started ☒

Tutorials

Documentation ☒

Developer Resources ☒

AWS Developer Forum ☒

Contact us ☒

Okay, here we are looking at the SageMaker console.

aws   Services   Q Search for services, features, blogs, docs, and more   [Option+S]   Oregon ▼   machine-learning ▼

Resource Groups & Tag Editor   S3   CloudFormation

Studio
Studio Lab
Canvas
RStudio

SageMaker dashboard
   Images
   Lifecycle configurations
   Search

▶ Ground Truth

▼ Notebook
   Notebook instances
   Git repositories

▶ Processing

▶ Training

▼ Inference
   Compilation jobs
   Marketplace model packages
   Models
   Endpoint configurations
   Endpoints
   Batch transform jobs

▶ Edge Manager

▶ Augmented AI

Max concurrent transforms	Max payload size (MB)	Batch strategy
-	-	-

Max invocation retries	Invocation timeout in seconds
-	-

Input data configuration

S3 data type	Split type	Compression type
S3Prefix	None	None

S3 URI	Content type
s3://rrs-machinelearning-media/input-images/	image/png

Output data configuration

S3 output path	Output data encryption key	Accept
s3://rrs-machinelearning-media/output-batch-1661921049/	-	-

The input data is the actual video frames

aws ⋮⋮⋮ Services   🔍 Search for services, features, blogs, docs, and more   [Option+S]   ⌨ ⊡   ⌃ △   ⊙   Oregon ▾   machine-learning ▾

🔲 Resource Groups & Tag Editor   🟩 S3   🔶 CloudFormation

Studio
Studio Lab 🔗   Amazon SageMaker > Batch transform jobs > batch-1661921049
Canvas
RStudio   # batch-1661921049   [Stop]

SageMaker dashboard   ## Job summary
Images
Lifecycle configurations   Job name   Status   Approx. batch transform duration
Search   batch-1661921049   ⊖ InProgress   -

▶ Ground Truth   Job ARN   Creation time
                                arn:aws:sagemaker:us-west-2:337745687079:transform-   Aug 31, 2022 04:44 UTC
▼ Notebook   job/batch-1661921049

Notebook instances
Git repositories   _____

▶ Processing   [Settings]   Monitor | Tags

▶ Training

▼ Inference   ## Job configuration

Compilation jobs
Marketplace model packages   Model name   Instance type   Instance count
Models   russ-gluoncv-yolo3-model   mL.m4.xlarge   1
Endpoint configurations
Endpoints   Encryption key
Batch transform jobs   -

▶ Edge Manager

▶ Augmented AI   Max concurrent transforms   Max payload size (MB)   Batch strategy
                                -   -

● ● ●

Amazon SageMaker � × | 🟧 rrs-machinelearning-media - ... × | ⬡ Home × | 🔳 batch_process_video_objects × | 🔳 russ-video-object-dect × | +

← → C 🔒 us-west-2.console.aws.amazon.com/s3/buckets/rrs-machinelearning-media?region=us-west-2&prefix=output-batch-1661921049/

aws ⠿ Services 🔍 Search for services, features, blogs, docs, and more [Option+S]   ⟩ ⏚ ⑦   Global ▼   machine-learning ▼

🔴 Resource Groups & Tag Editor   🟩 S3   🟥 CloudFormation

Amazon S3   ✕

Buckets

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

Access analyzer for S3

Block Public Access settings for
this account

▼ **Storage Lens**

Dashboards

AWS Organizations settings

Feature spotlight 3

▶ AWS Marketplace for S3

Amazon S3 > Buckets > rrs-machinelearning-media > output-batch-1661921049/

[🗗 Copy S3 URI]

output-batch-1661921049/

Objects	Properties

Objects (5)

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory ☑ to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more ☑

[↻] [🗗 Copy S3 URI] [🗗 Copy URL] [⬇ Download] [Open ↗] [Delete] [Actions ▼] [Create folder] [⬆ Upload]

🔍 Find objects by prefix   ‹ 1 › ⚙

☐	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class ▽
☐	🗋 output_0001.png.out	out	August 31, 2022, 14:49:03 (UTC+10:00)	103.0 B	Standard
☐	🗋 output_0002.png.out	out	August 31, 2022, 14:48:57 (UTC+10:00)	103.0 B	Standard
☐	🗋 output_0003.png.out	out	August 31, 2022, 14:48:59 (UTC+10:00)	312.0 B	Standard
☐	🗋 output_0004.png.out	out	August 31, 2022, 14:49:04 (UTC+10:00)	206.0 B	Standard
☐	🗋 output_0005.png.out	out	August 31, 2022, 14:49:01 (UTC+10:00)	211.0 B	Standard

the transform operation,