Sagemaker

Session-5

CLARUSWAY©
WAY TO REINVENT YOURSELF

Table of Contents

- Cloud Computing Market Size
- What is SageMaker?
- SageMaker Process
- Algorithms
- Notebook Instance
- S3
- Billing

I’ve completed the pre-class content?

Yes

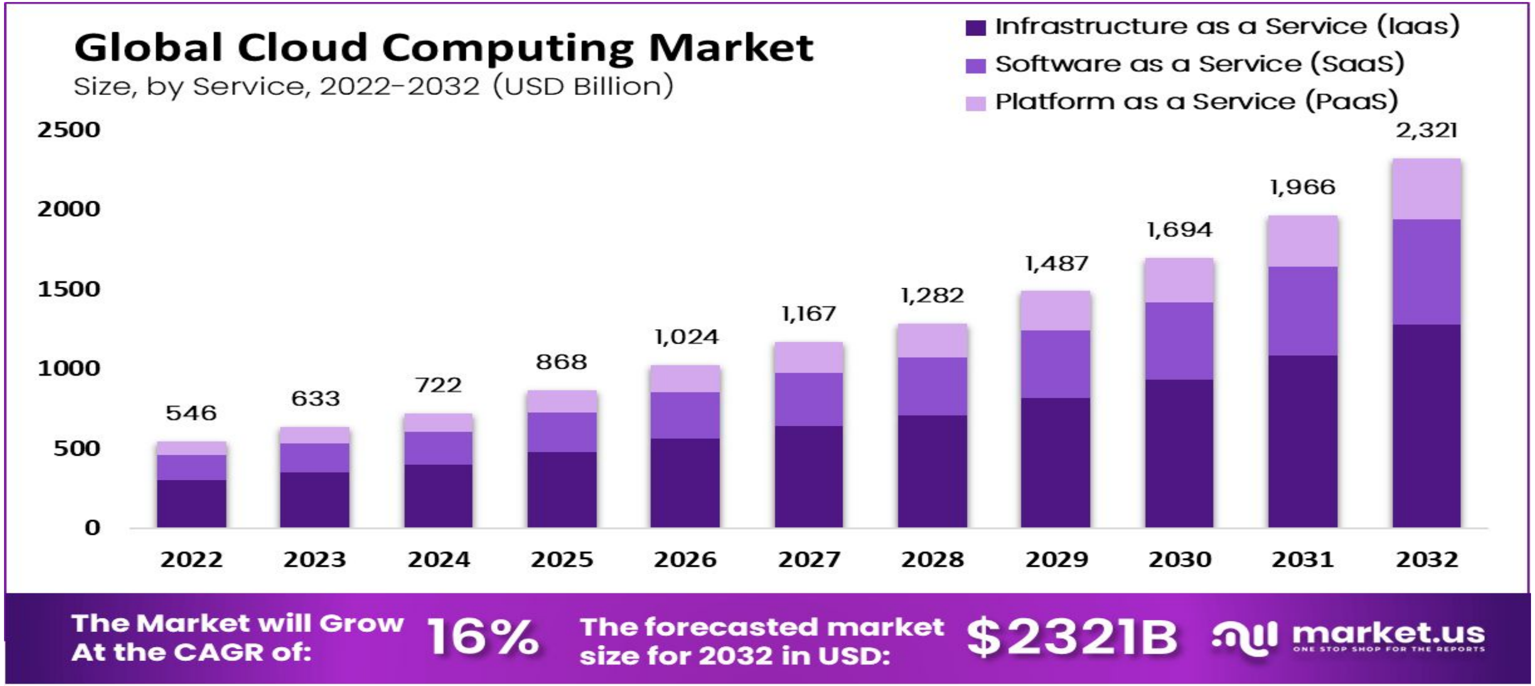
No

Students choose an option

SageMaker

CLARUSWAY©
WAY TO REINVENT YOURSELF

Cloud Computing Market Size



Global Cloud Computing Market

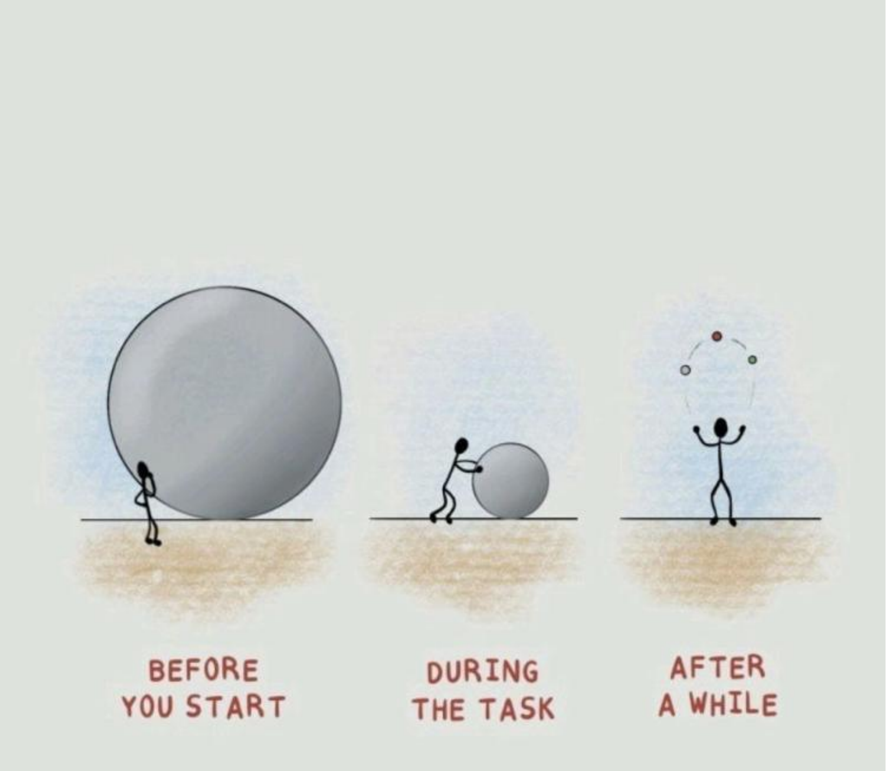
Size, by Service, 2022-2032 (USD Billion)

The Market will Grow At the CAGR of: 16%

The forecasted market size for 2032 in USD: $2321B

- The rising popularity of the latest novel technologies like **artificial intelligence** and **machine learning** and its rapid adoption in the cloud computing is empowering the growth of the global cloud computing market.

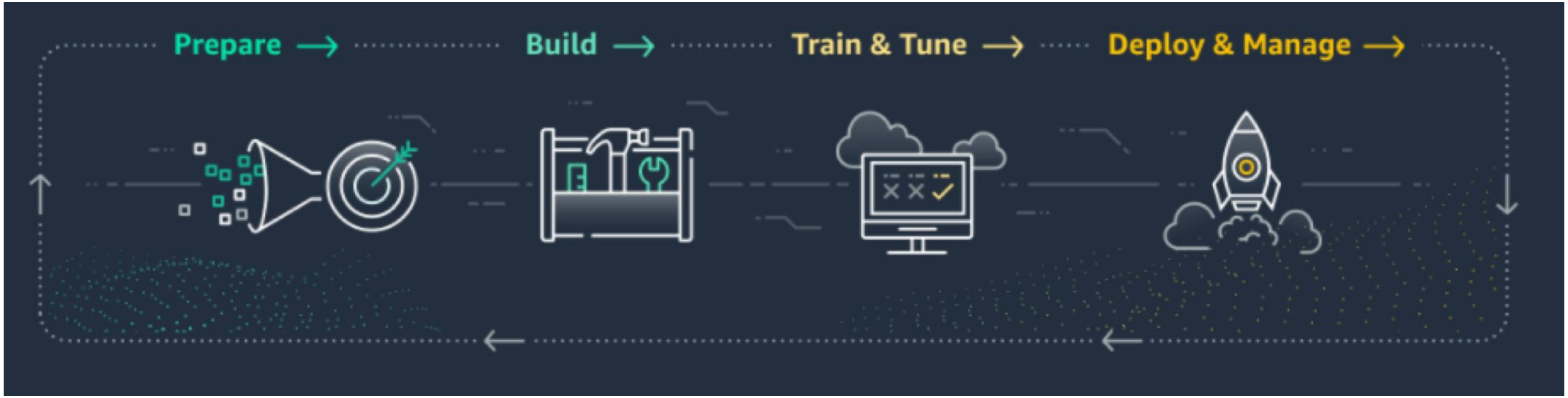
SageMaker



Alex Maese Juarez

What is SageMaker

Amazon **SageMaker** is a **platform** service that simplifies the process of **building, training, and deploying ML models** by providing everything organizations need to connect to their training data, select and optimize the **best algorithm and framework**, and **deploy their model** on auto-scaling clusters of Amazon **EC2**.



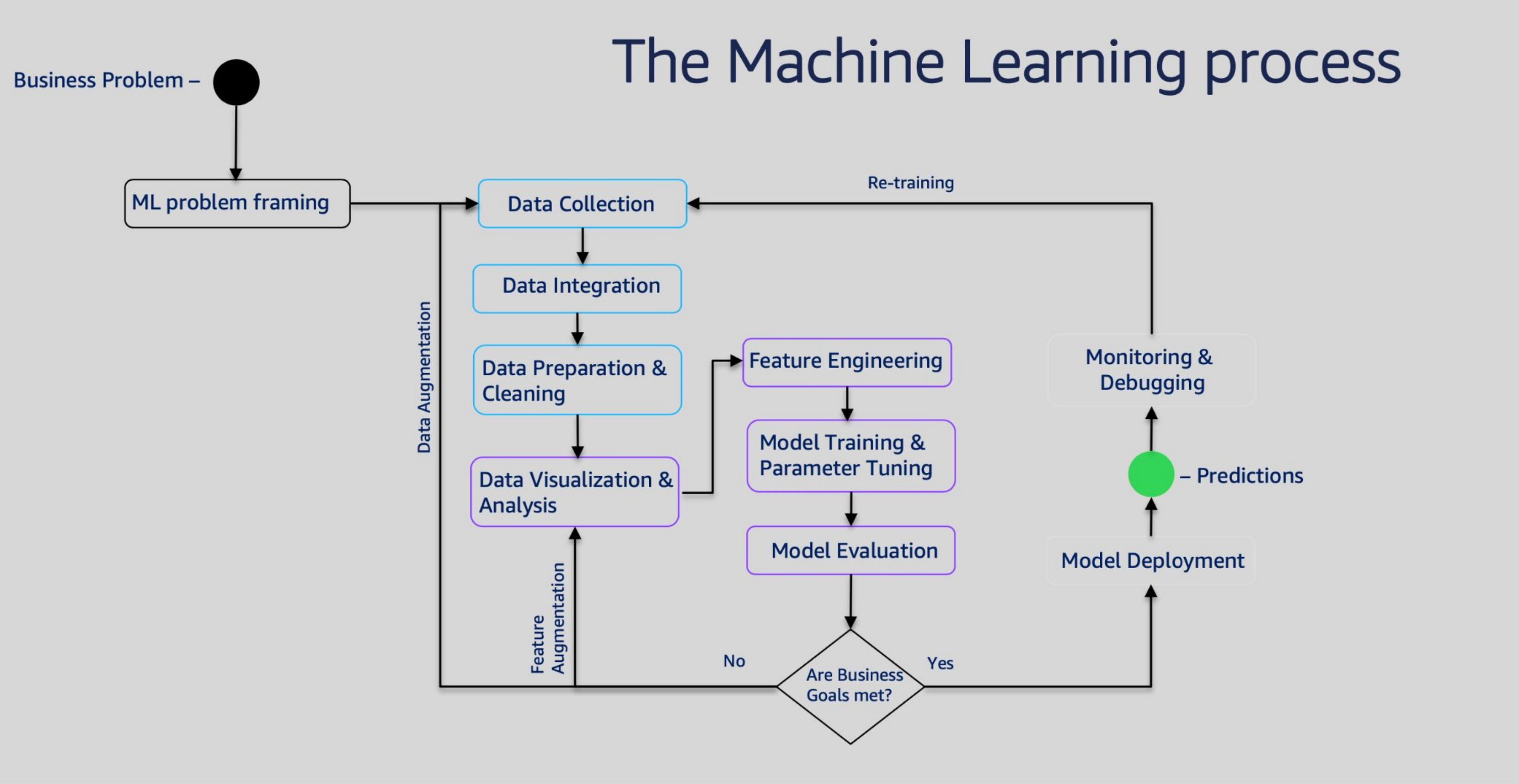
WHY Amazon SageMaker?

- Accelerating machine learning innovation through security
- Security features from Amazon SageMaker and the AWS Cloud can help organizations go from idea to production faster
- https://amer.resources.awscloud.com/ai-ml/accelerating-machine-learning-innovation-through-security

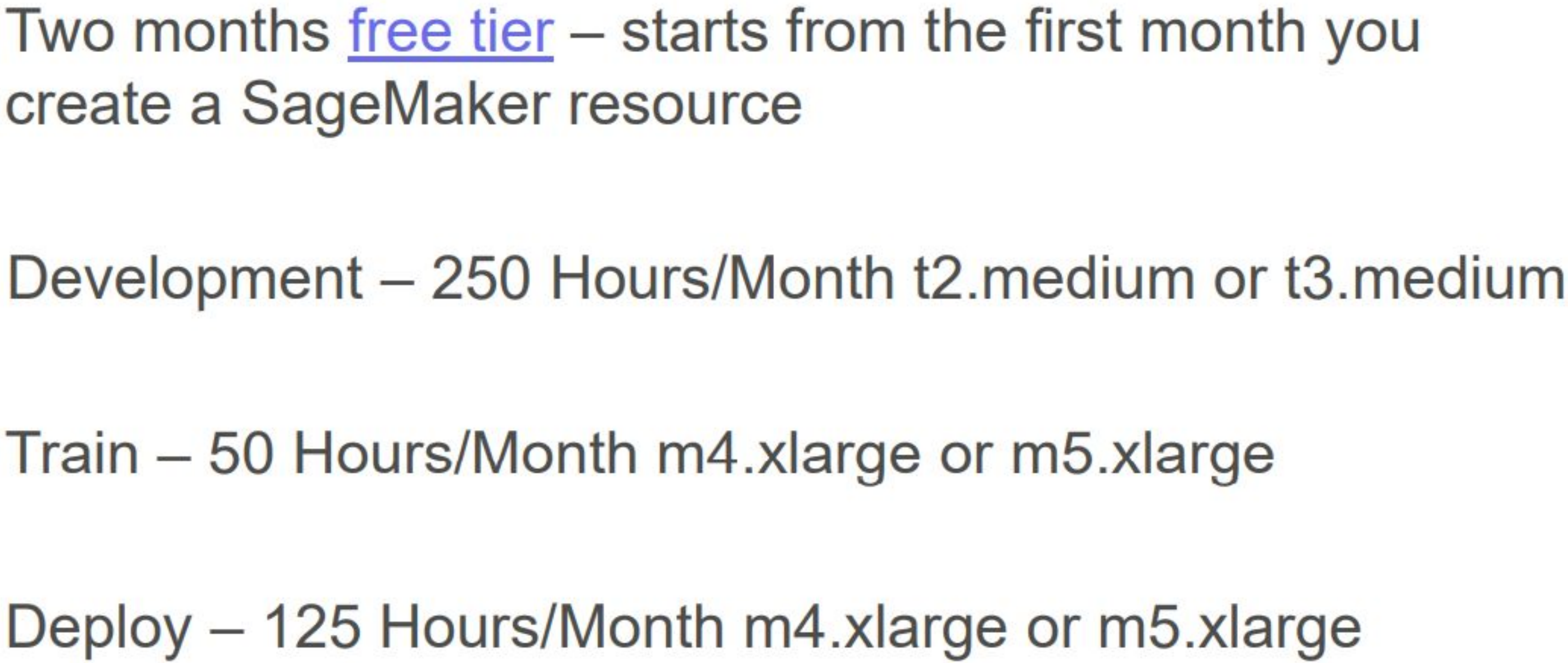
ML with AWS, by the numbers

- Reduce training time by 50%
- Provide 90% scaling efficiency
- Deliver 3x faster network throughput
- Improve price and performance by 25%
- 91% of cloud-based PyTorch runs on AWS
- 92% of cloud-based TensorFlow runs on AWS

ML process


The Machine Learning process
Business Problem –
ML problem framing
Data Collection
Re-training
Data Integration
Data Augmentation
Data Preparation & Cleaning
Feature Engineering
Monitoring & Debugging
Model Training & Parameter Tuning
– Predictions
Data Visualization & Analysis
Model Evaluation
Model Deployment
Feature Augmentation
No
Are Business Goals met?
Yes

SageMaker Free Tier (2 months)



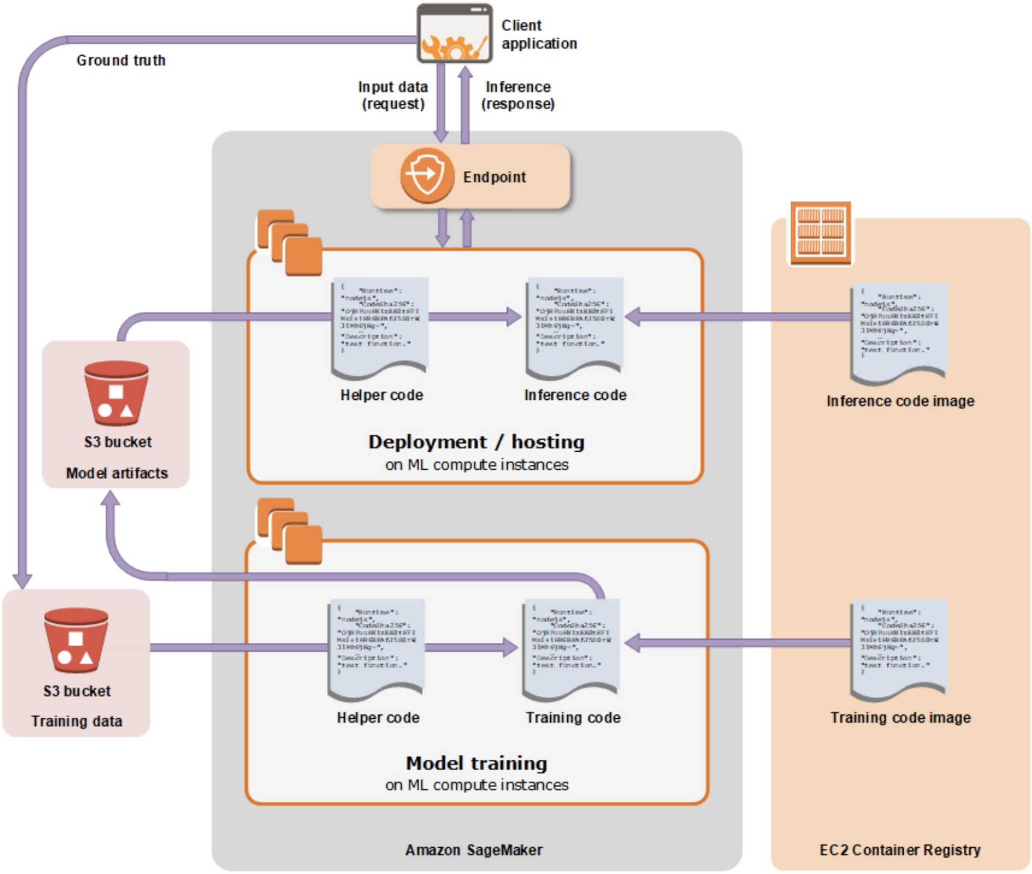
Two months free tier – starts from the first month you create a SageMaker resource

Development – 250 Hours/Month t2.medium or t3.medium

Train – 50 Hours/Month m4.xlarge or m5.xlarge

Deploy – 125 Hours/Month m4.xlarge or m5.xlarge

Model Deployment Architecture



SageMaker Process

- Get your data

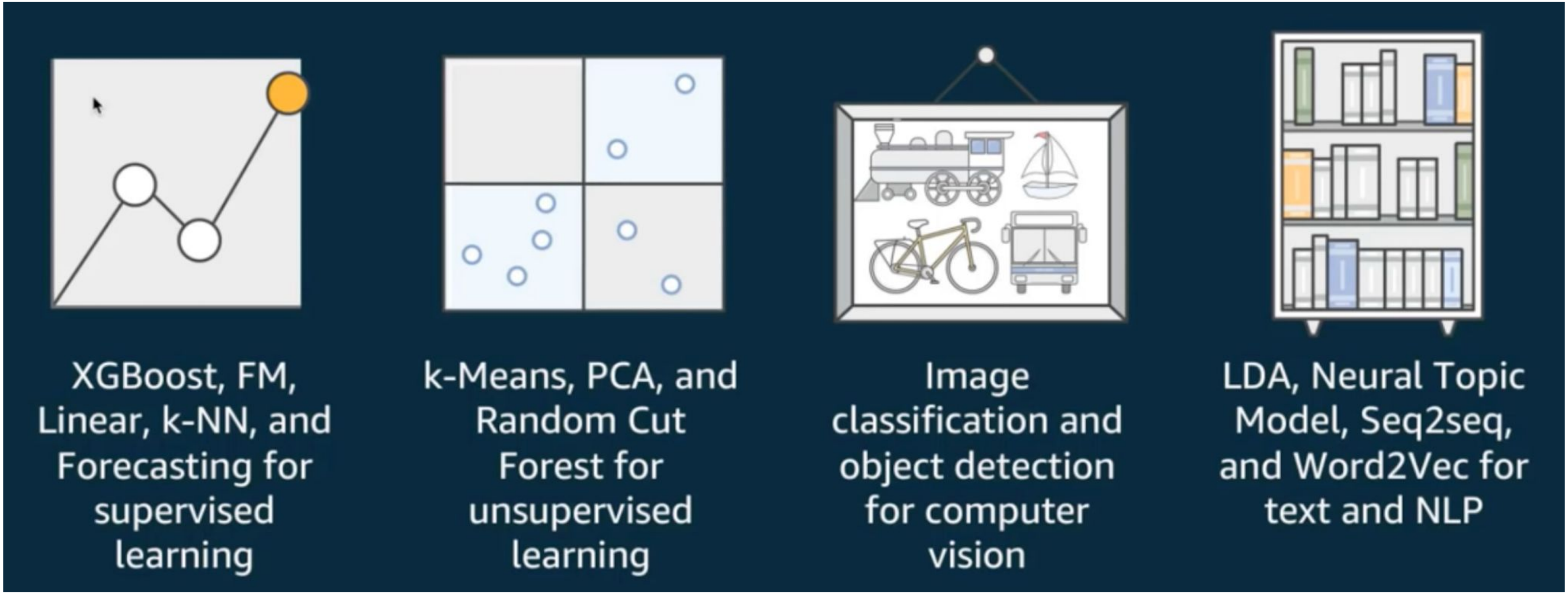
XG-Boost
TensorFlow™
mxnet
• Linear Learner

- Explore and refine models in a single Notebook Instance

- Train on the full dataset in a cluster of GPU instances..

- Deploy to production

Built-in algorithms



XGBoost, FM, Linear, k-NN, and Forecasting for supervised learning

k-Means, PCA, and Random Cut Forest for unsupervised learning

Image classification and object detection for computer vision

LDA, Neural Topic Model, Seq2seq, and Word2Vec for text and NLP

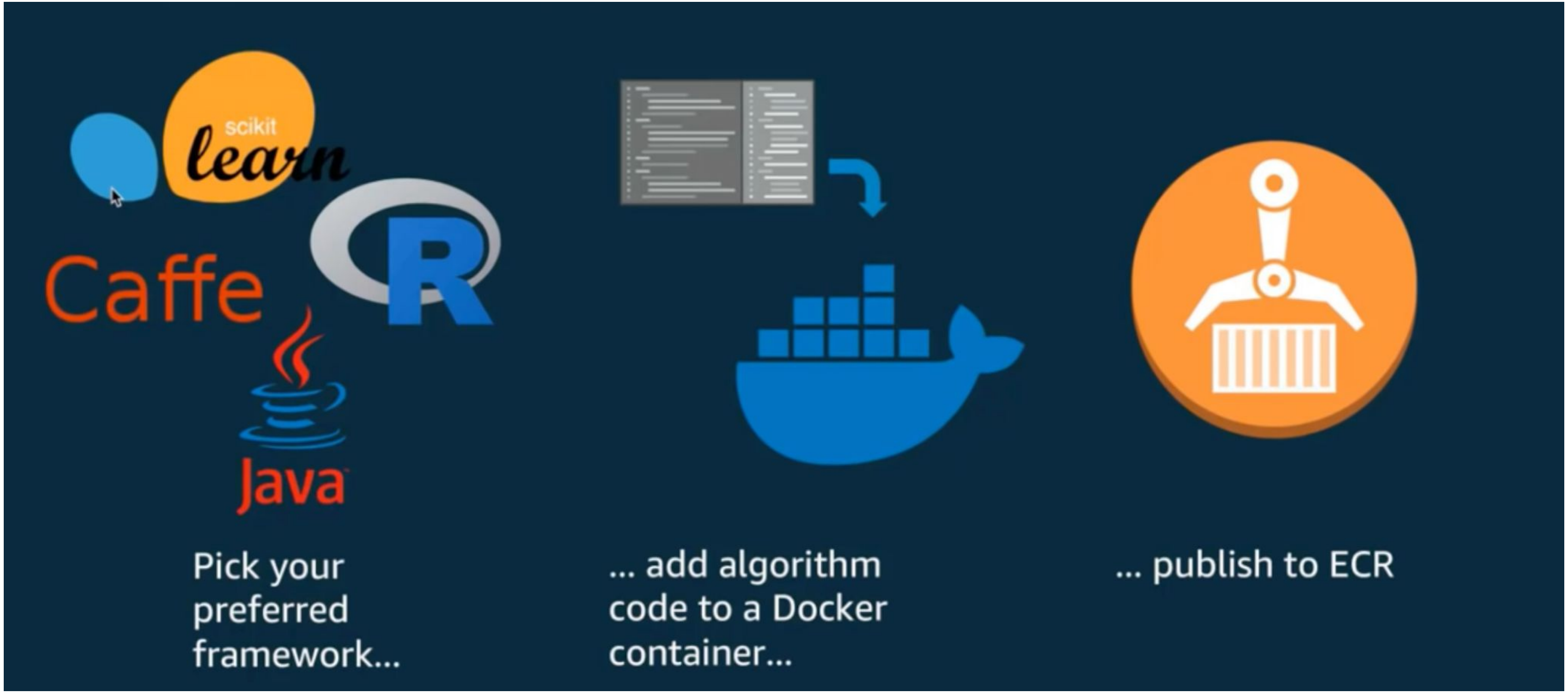
Built-in algorithms-Supervised

- Linear Learner: regression
- K-Nearest Neighbors: non-parametric regression and classification
- XGBoost: regression, classification
- Factorization Machines: regression, classification, recommendation
- Semantic Segmentation: Deep Learning
- Image Classification: Deep Learning (ResNet)
- Object Detection (SSD): Deep Learning
- (VGG or ResNet)
- Sequence to Sequence: machine translation, speech to text and more
- DeepAR: time-series forecasting (RNN)

Built-in algorithms-Unsupervised

- K-Means: clustering
- Principal Component Analysis: dimensionality reduction
- Random Cut Forest: anomaly detection
- Object2Vec: general-purpose embedding
- Neural Topic Model: topic modeling
- Latent Dirichlet Allocation: topic modeling (mostly)
- Blazing Text: GPU-based Word2 Vec, and text classification
- IP Insights: usage patterns for IP addresses

Bring your own algorithm



Pick your preferred framework...

... add algorithm code to a Docker container...

... publish to ECR

Let’s jump to the AWS SageMaker Service!

S3 BUCKET and OBJECTS

- iden-ml-sagemaker/
  - bikerental/
    - train/
      - train.csv
    - validation/
      - validation.csv
    - test/
      - test.csv
    - model/
      - xgboost-bikerental-v1-2022-04-21-10-23-10-964/ (Job_name+datetime)
        - output
          - model.tar.gz

Amazon S3>Buckets>iden-ml-sagemaker>bikerental/output/xgbobikerental-v1-2022-04-21-10-23-10-964/output/model.tar.gz

Notebook Instance

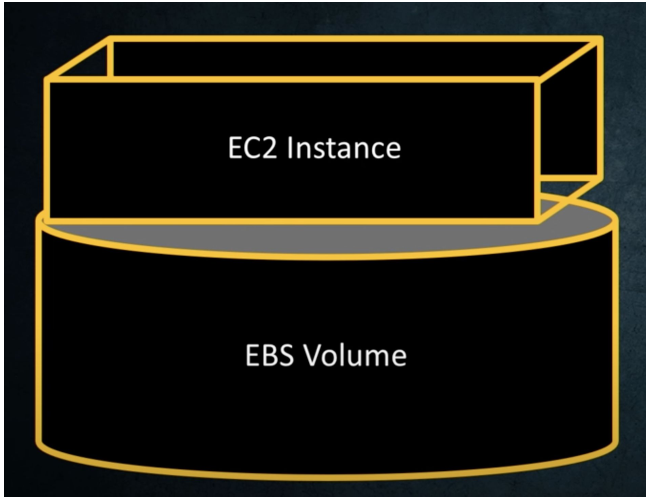
An *Amazon SageMaker notebook instance* is a machine learning (ML) compute instance running the Jupyter Notebook App. SageMaker manages creating the instance and related resources.

Notebook Instance

Use Jupyter notebooks in your notebook instance

- To prepare and process data,
- Write code to train models,
- Deploy models to SageMaker hosting,
- Test or validate your models.

Notebook Instance



- Anaconda Packages,
- Tensorflow and Apache MXnet,
- Storage volume,
- Sample notebooks that contain complete code walkthroughs

You are using CSV formatted files to train on SageMaker's built-in XGBoost algorithm. SageMaker expects your training and validation to follow this convention:

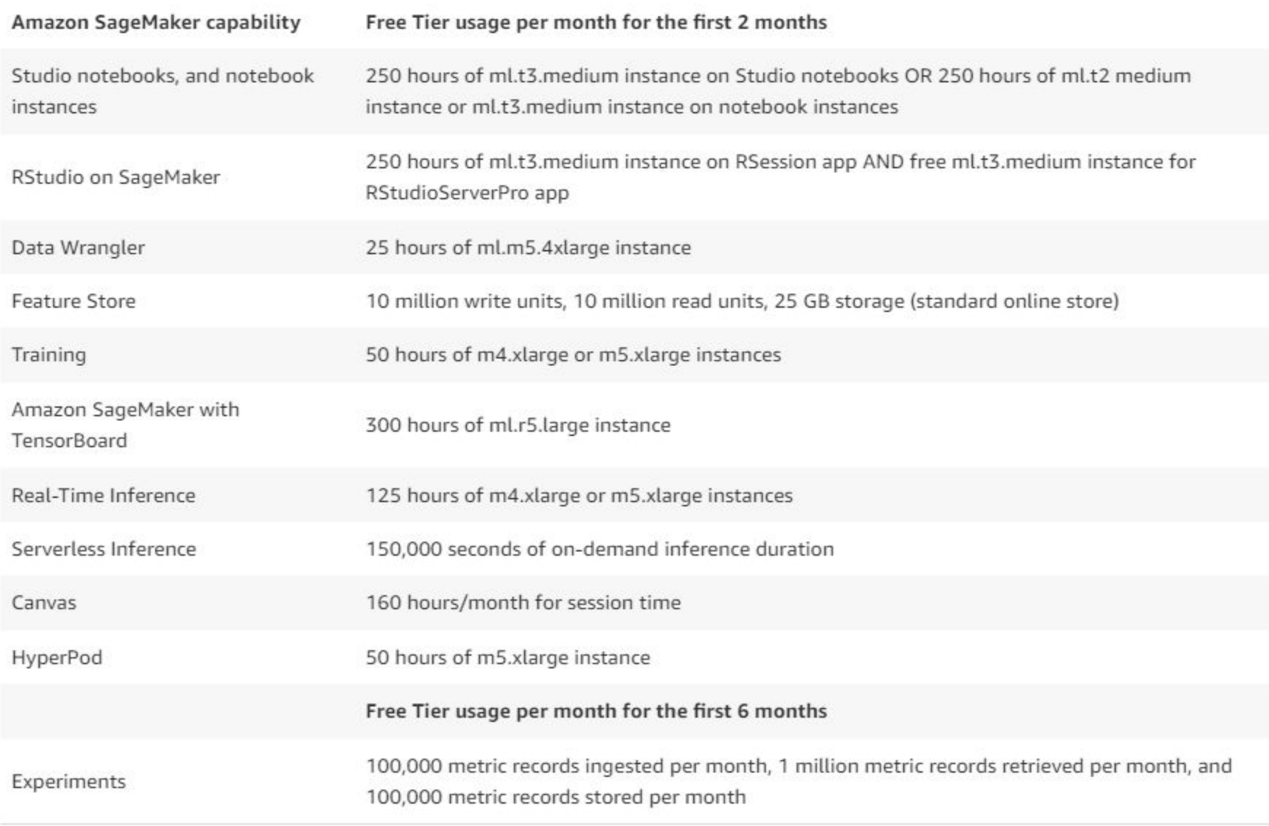
- CSV must have column headers and target variable must be the last column
- CSV must have column headers with the target variable in the first column
- CSV must not have a column header record. Target variable must be the last column
- CSV must not have a column header record. Target variable must be the first column

How does SageMaker built-in know the target variable?

- For CSV training, the algorithm assumes that the target variable is in the first column and that the CSV does not have a header record.(Train,Validation)
- For CSV test, the algorithm assumes that CSV input does not have the label column.

https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost.html

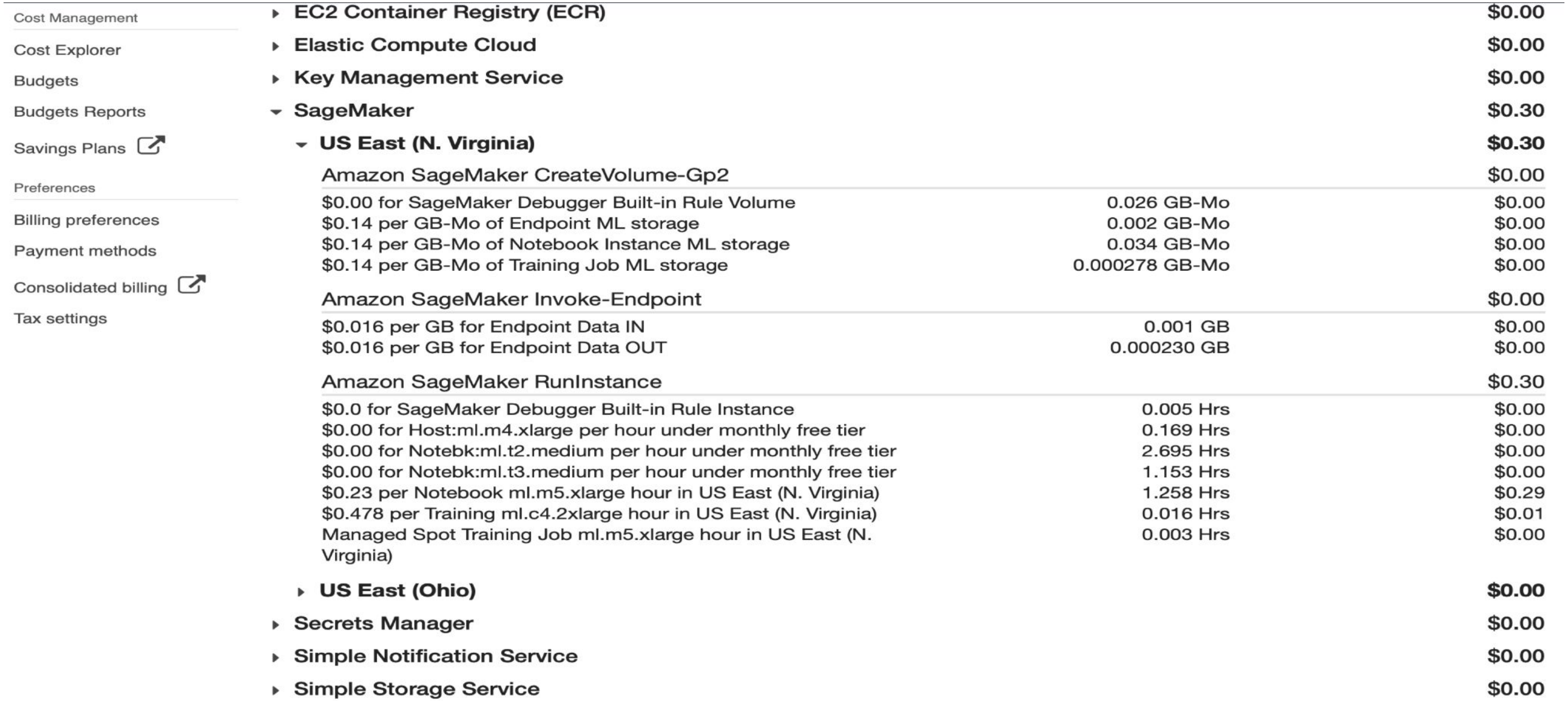
SageMaker Free Tier Usage



Amazon SageMaker capability	**Free Tier usage per month for the first 2 months**
Studio notebooks, and notebook instances	250 hours of ml.t3.medium instance on Studio notebooks OR 250 hours of ml.t2 medium instance or ml.t3.medium instance on notebook instances
RStudio on SageMaker	250 hours of ml.t3.medium instance on RSession app AND free ml.t3.medium instance for RStudioServerPro app
Data Wrangler	25 hours of ml.m5.4xlarge instance
Feature Store	10 million write units, 10 million read units, 25 GB storage (standard online store)
Training	50 hours of m4.xlarge or m5.xlarge instances
Amazon SageMaker with TensorBoard	300 hours of ml.r5.large instance
Real-Time Inference	125 hours of m4.xlarge or m5.xlarge instances
Serverless Inference	150,000 seconds of on-demand inference duration
Canvas	160 hours/month for session time
HyperPod	50 hours of m5.xlarge instance
	Free Tier usage per month for the first 6 months
Experiments	100,000 metric records ingested per month, 1 million metric records retrieved per month, and 100,000 metric records stored per month

https://aws.amazon.com/sagemaker/pricing/

SageMaker Billing Dashboard



Cost Management

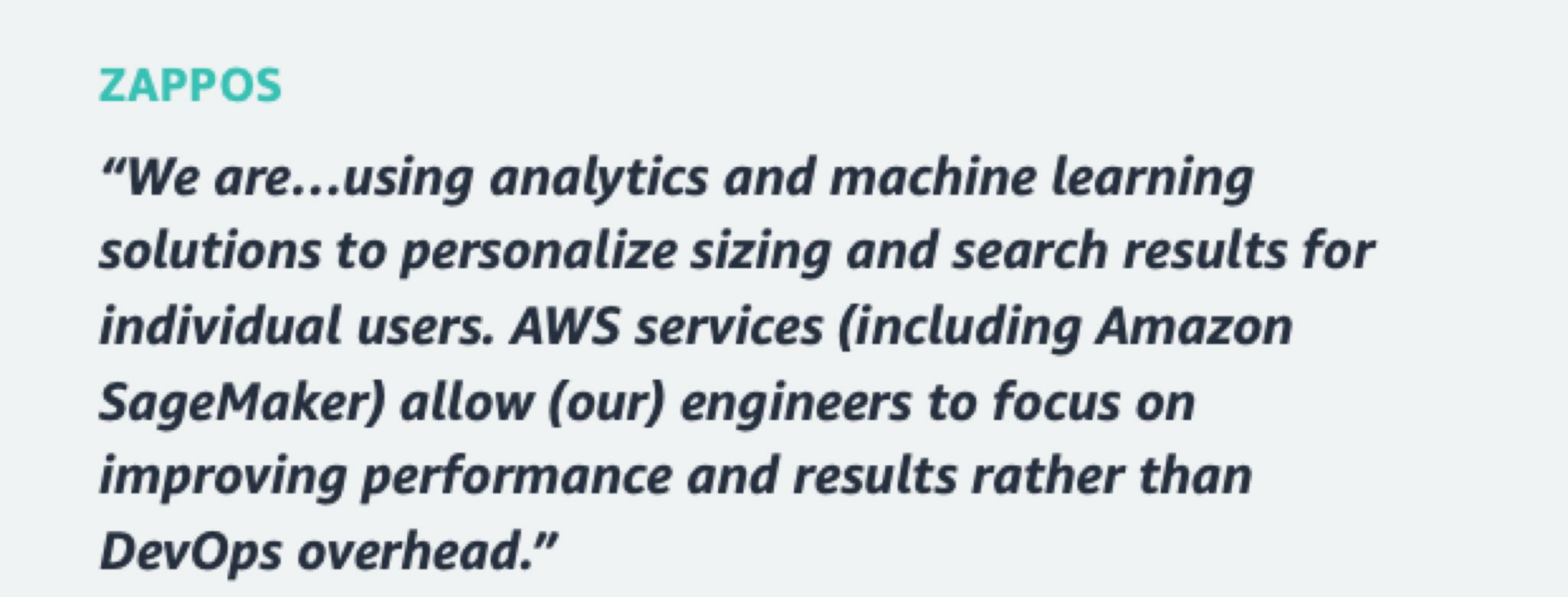
- Cost Explorer
- Budgets
- Budgets Reports
- Savings Plans

Preferences

- Billing preferences
- Payment methods
- Consolidated billing
- Tax settings

▸ **EC2 Container Registry (ECR)**		**$0.00**
▸ **Elastic Compute Cloud**		**$0.00**
▸ **Key Management Service**		**$0.00**
▾ **SageMaker**		**$0.30**
▾ **US East (N. Virginia)**		**$0.30**
Amazon SageMaker CreateVolume-Gp2		$0.00
$0.00 for SageMaker Debugger Built-in Rule Volume	0.026 GB-Mo	$0.00
$0.14 per GB-Mo of Endpoint ML storage	0.002 GB-Mo	$0.00
$0.14 per GB-Mo of Notebook Instance ML storage	0.034 GB-Mo	$0.00
$0.14 per GB-Mo of Training Job ML storage	0.000278 GB-Mo	$0.00
Amazon SageMaker Invoke-Endpoint		$0.00
$0.016 per GB for Endpoint Data IN	0.001 GB	$0.00
$0.016 per GB for Endpoint Data OUT	0.000230 GB	$0.00
Amazon SageMaker RunInstance		$0.30
$0.0 for SageMaker Debugger Built-in Rule Instance	0.005 Hrs	$0.00
$0.00 for Host:ml.m4.xlarge per hour under monthly free tier	0.169 Hrs	$0.00
$0.00 for Notebk:ml.t2.medium per hour under monthly free tier	2.695 Hrs	$0.00
$0.00 for Notebk:ml.t3.medium per hour under monthly free tier	1.153 Hrs	$0.00
$0.23 per Notebook ml.m5.xlarge hour in US East (N. Virginia)	1.258 Hrs	$0.29
$0.478 per Training ml.c4.2xlarge hour in US East (N. Virginia)	0.016 Hrs	$0.01
Managed Spot Training Job ml.m5.xlarge hour in US East (N. Virginia)	0.003 Hrs	$0.00
▸ **US East (Ohio)**		**$0.00**
▸ **Secrets Manager**		**$0.00**
▸ **Simple Notification Service**		**$0.00**
▸ **Simple Storage Service**		**$0.00**

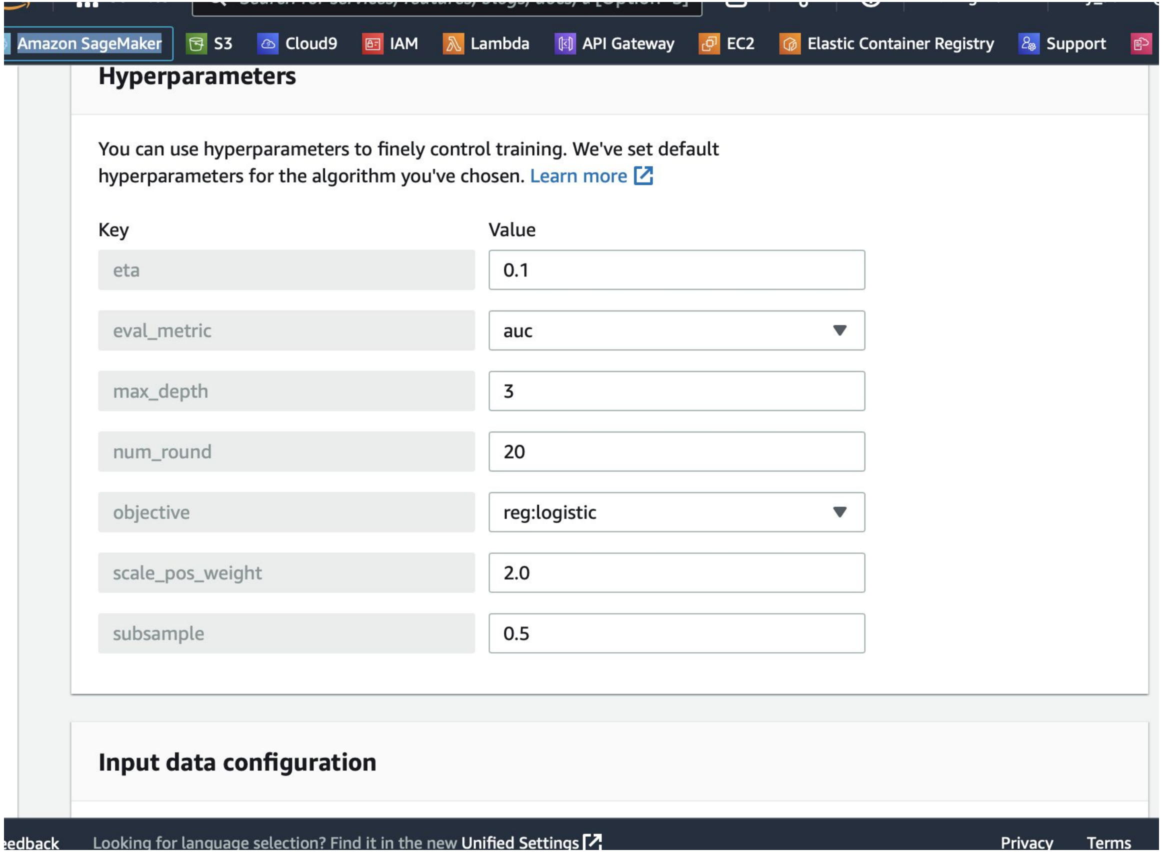
SageMaker Use Case Example



ZAPPOS

"We are...using analytics and machine learning solutions to personalize sizing and search results for individual users. AWS services (including Amazon SageMaker) allow (our) engineers to focus on improving performance and results rather than DevOps overhead."

Hyperparameters



Hyperparameters

You can use hyperparameters to finely control training. We've set default hyperparameters for the algorithm you've chosen. Learn more

Key	Value
eta	0.1
eval_metric	auc ▼
max_depth	3
num_round	20
objective	reg:logistic ▼
scale_pos_weight	2.0
subsample	0.5

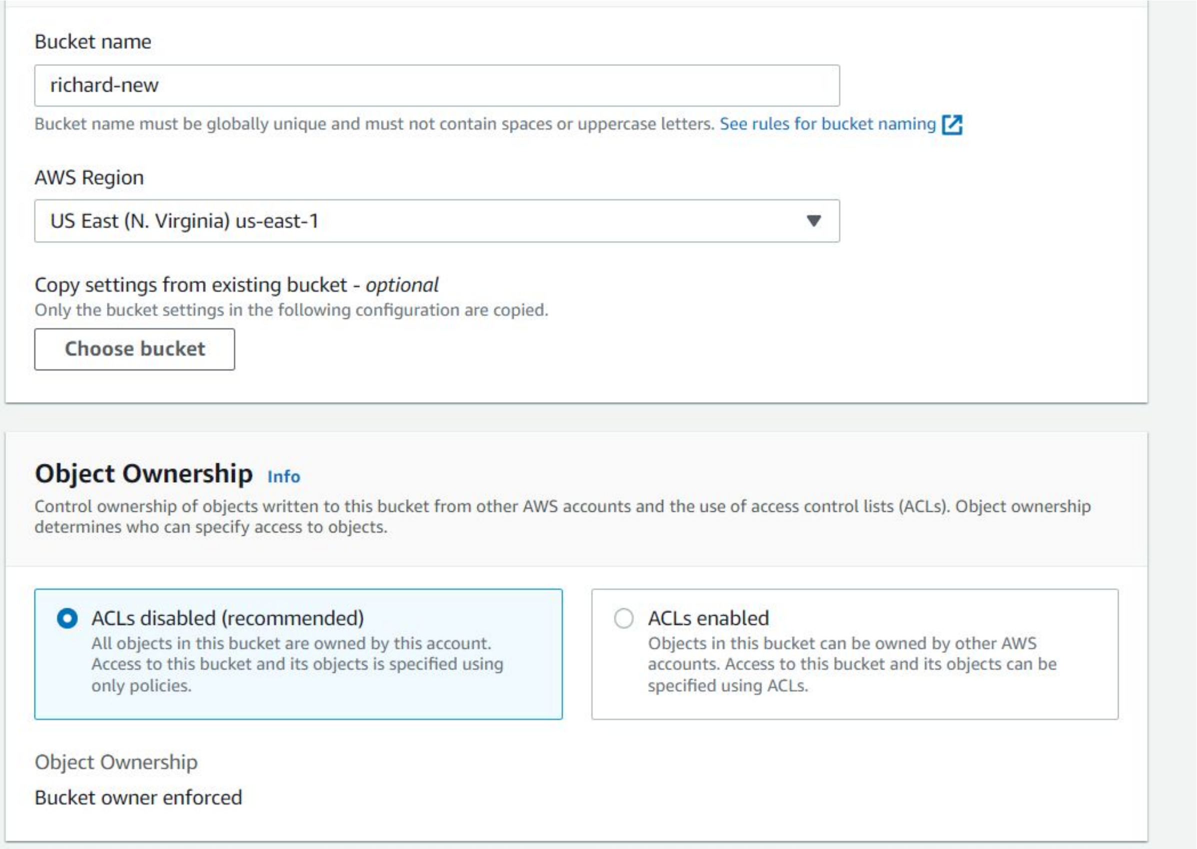
Input data configuration

Looking for language selection? Find it in the new Unified Settings

THANKS!

Any questions?

S3 Settings



Bucket name

richard-new

Bucket name must be globally unique and must not contain spaces or uppercase letters. See rules for bucket naming

AWS Region

US East (N. Virginia) us-east-1 ▼

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

Object Ownership

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

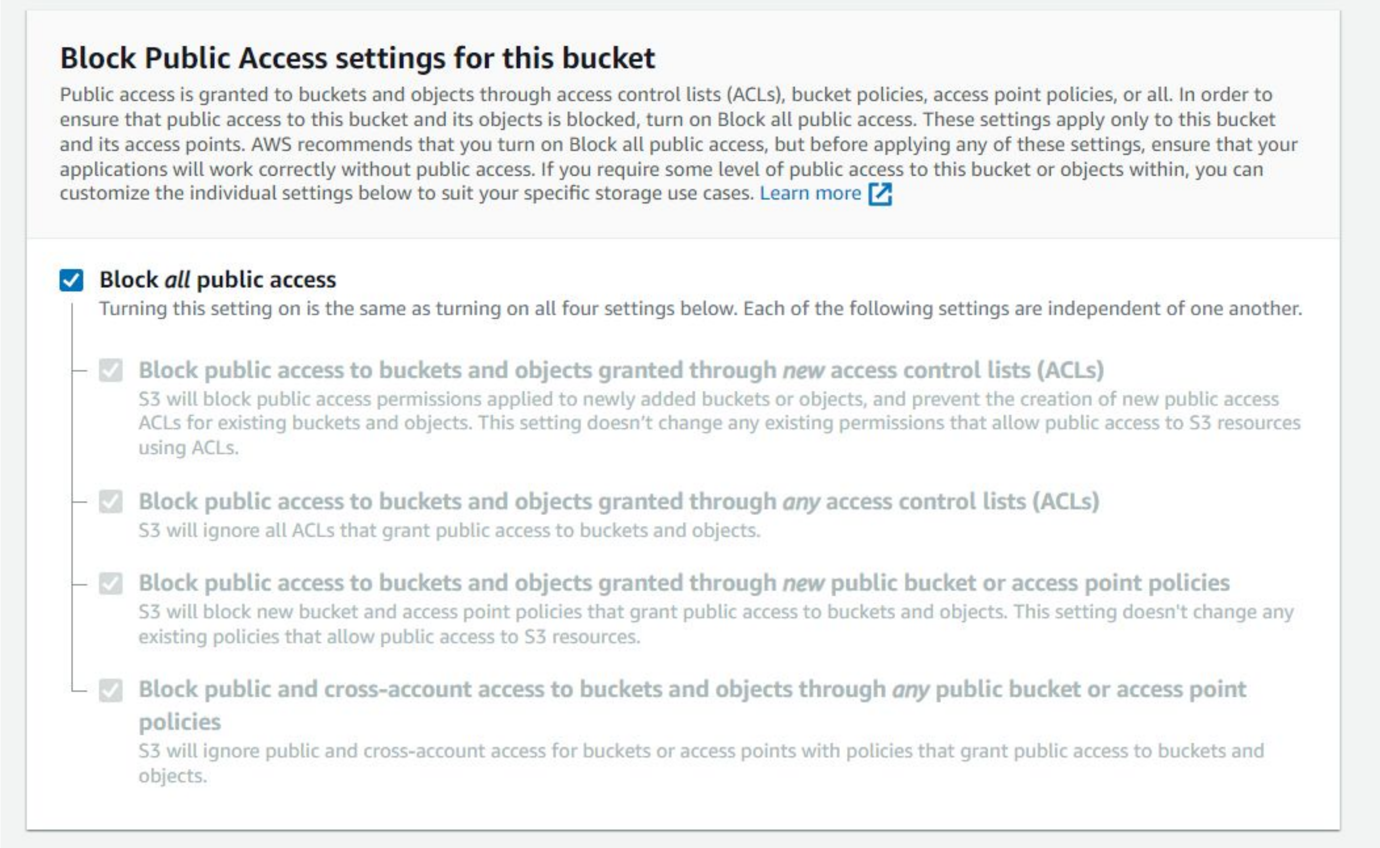
● ACLs disabled (recommended)
All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.

○ ACLs enabled
Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

Object Ownership

Bucket owner enforced

S3 Settings



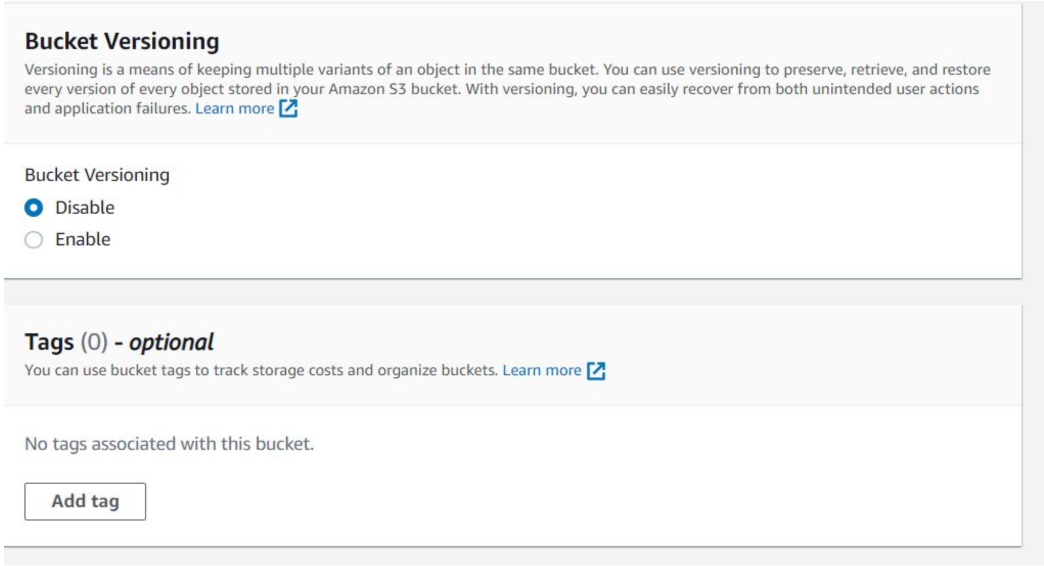
Block Public Access settings for this bucket

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. Learn more

☑ **Block *all* public access**
Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.

- ☑ **Block public access to buckets and objects granted through *new* access control lists (ACLs)**
  S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.
- ☑ **Block public access to buckets and objects granted through *any* access control lists (ACLs)**
  S3 will ignore all ACLs that grant public access to buckets and objects.
- ☑ **Block public access to buckets and objects granted through *new* public bucket or access point policies**
  S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.
- ☑ **Block public and cross-account access to buckets and objects through *any* public bucket or access point policies**
  S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

S3 Settings



Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. Learn more

Bucket Versioning

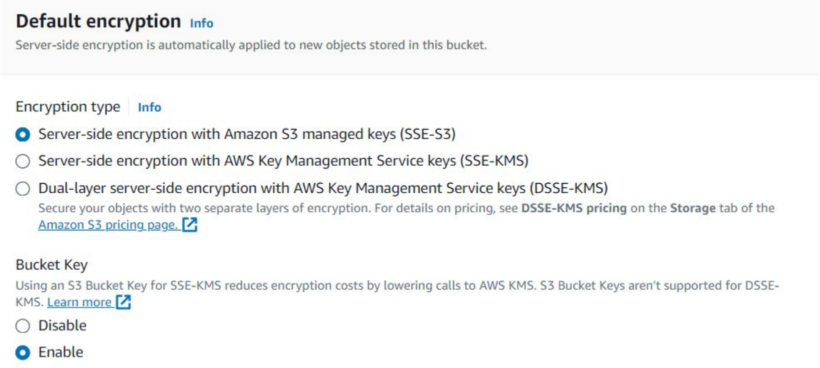
● Disable

○ Enable

Tags (0) - *optional*

You can use bucket tags to track storage costs and organize buckets. Learn more

No tags associated with this bucket.



Default encryption Info

Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type | Info

● Server-side encryption with Amazon S3 managed keys (SSE-S3)

○ Server-side encryption with AWS Key Management Service keys (SSE-KMS)

○ Dual-layer server-side encryption with AWS Key Management Service keys (DSSE-KMS)

Secure your objects with two separate layers of encryption. For details on pricing, see **DSSE-KMS pricing** on the **Storage** tab of the Amazon S3 pricing page.

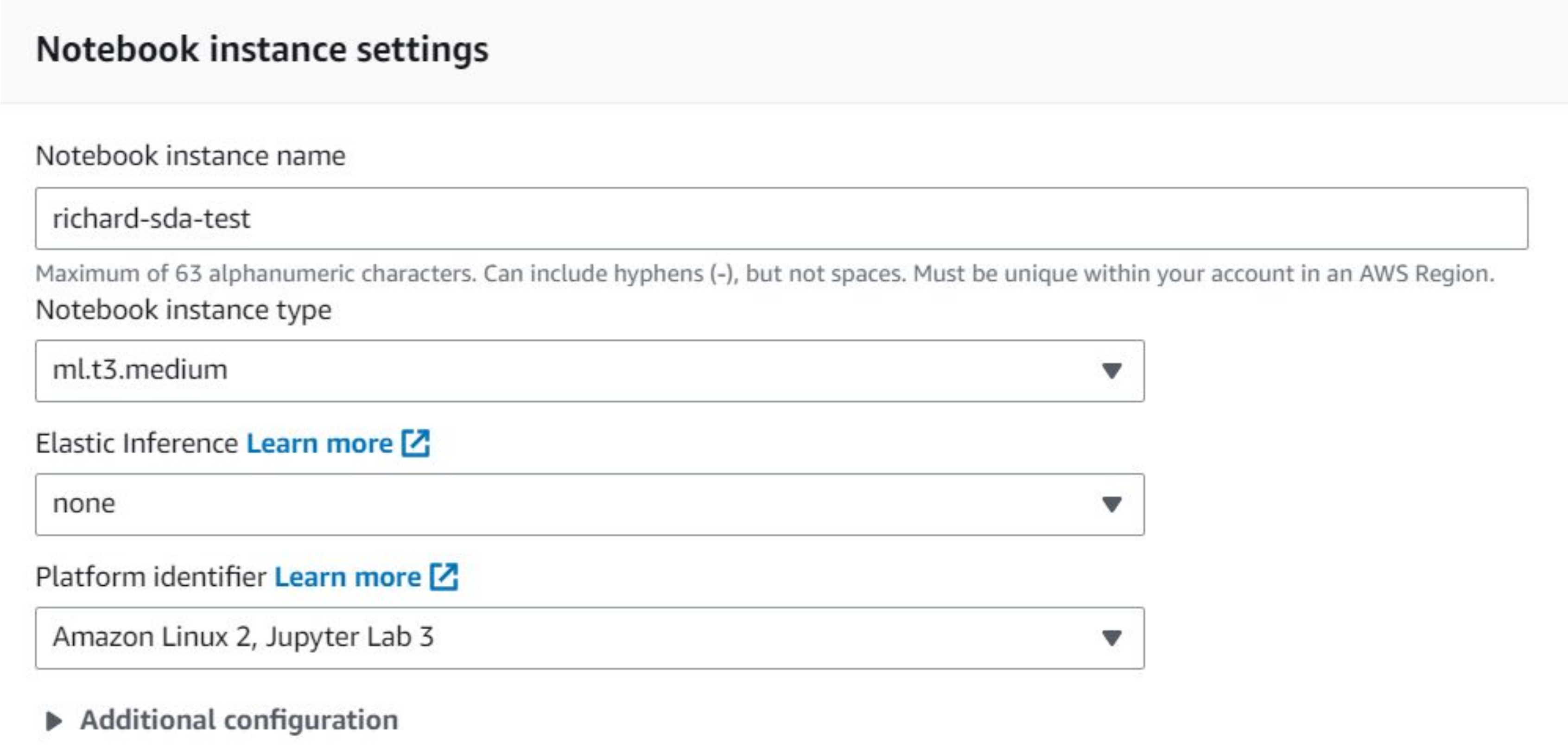
Bucket Key

Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS. Learn more

○ Disable

● Enable

Notebook Settings



Notebook instance settings

Notebook instance name

richard-sda-test

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.t3.medium ▼

Elastic Inference Learn more

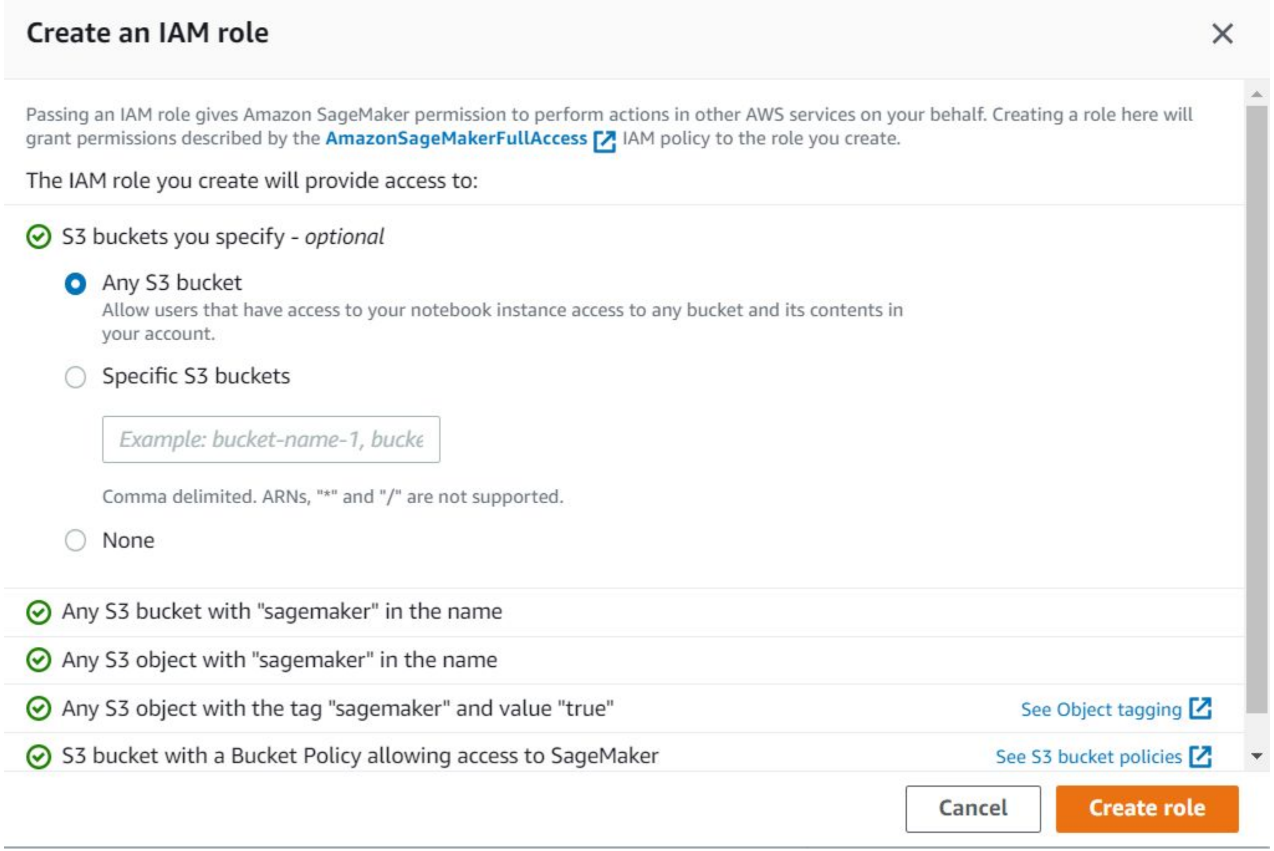
none ▼

Platform identifier Learn more

Amazon Linux 2, Jupyter Lab 3 ▼

▶ Additional configuration

Notebook Settings



Create an IAM role

Passing an IAM role gives Amazon SageMaker permission to perform actions in other AWS services on your behalf. Creating a role here will grant permissions described by the **AmazonSageMakerFullAccess** IAM policy to the role you create.

The IAM role you create will provide access to:

✓ S3 buckets you specify - *optional*

● Any S3 bucket
Allow users that have access to your notebook instance access to any bucket and its contents in your account.

○ Specific S3 buckets

Example: bucket-name-1, bucke

Comma delimited. ARNs, "*" and "/" are not supported.

○ None

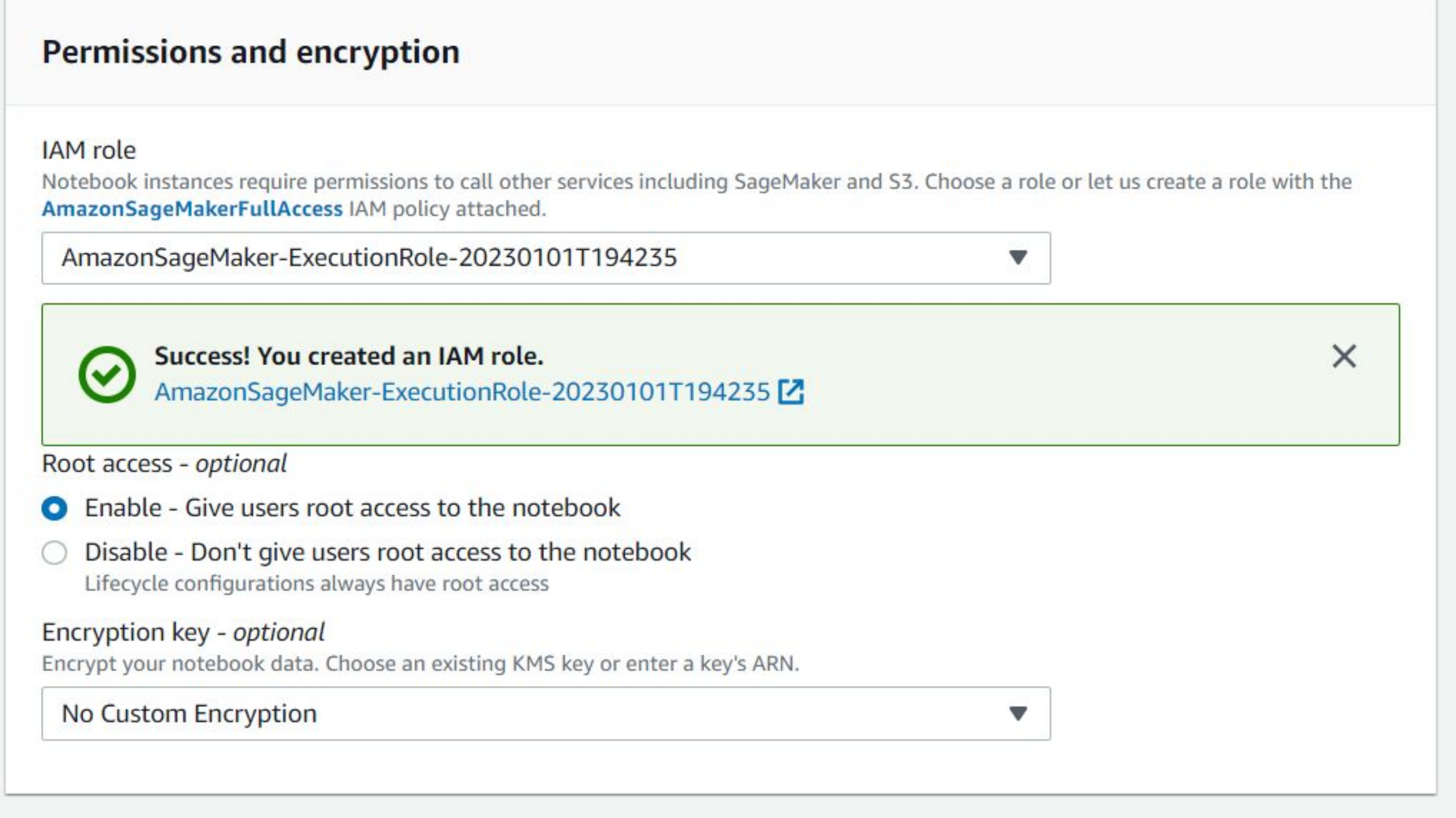
✓ Any S3 bucket with "sagemaker" in the name

✓ Any S3 object with "sagemaker" in the name

✓ Any S3 object with the tag "sagemaker" and value "true" — See Object tagging

✓ S3 bucket with a Bucket Policy allowing access to SageMaker — See S3 bucket policies

Notebook Settings



Permissions and encryption

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-20230101T194235 ▼

Success! You created an IAM role.
AmazonSageMaker-ExecutionRole-20230101T194235

Root access - *optional*

● Enable - Give users root access to the notebook

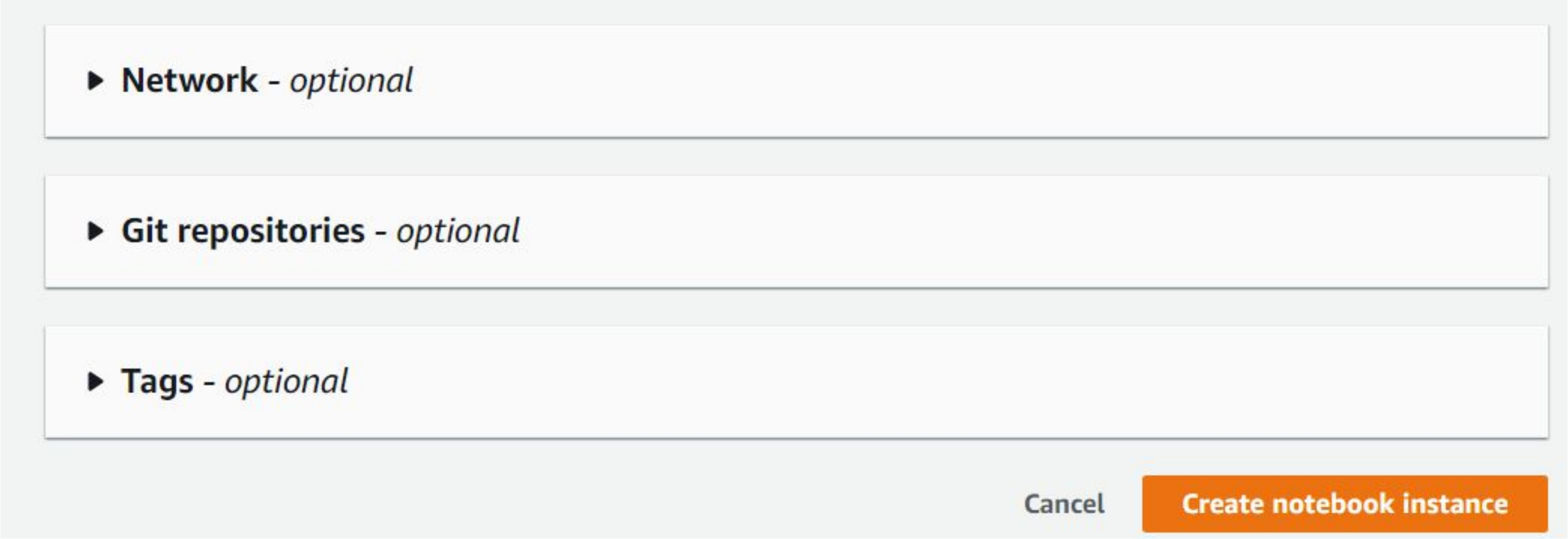
○ Disable - Don't give users root access to the notebook
Lifecycle configurations always have root access

Encryption key - *optional*

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

No Custom Encryption ▼

Notebook Settings



▸ **Network** - *optional*

▸ **Git repositories** - *optional*

▸ **Tags** - *optional*

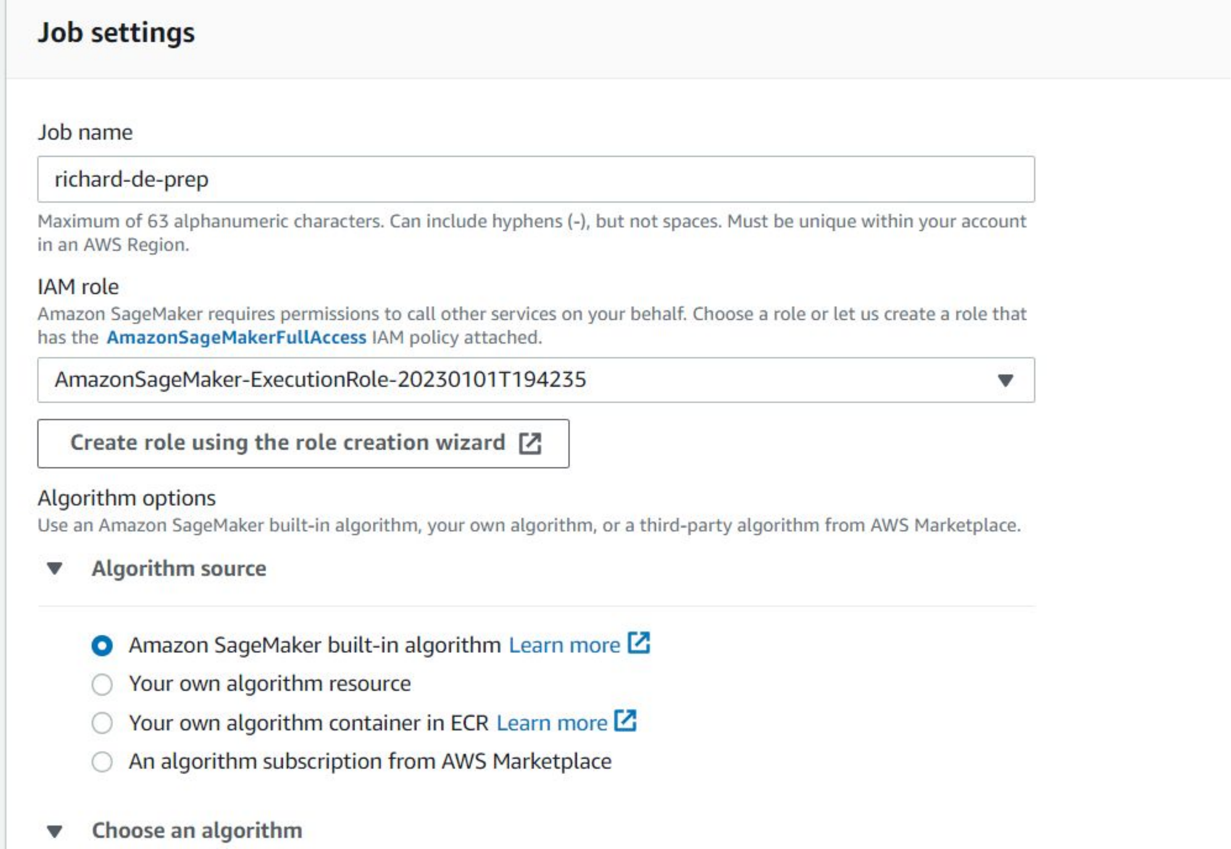
Notebook Settings

Now open the notebook from Sagemaker and send the data S3 via notebook

After sending data come back to the sagemaker console and do the jobs:

Training Job Creation-Model Creation- End Point Creation

Train Job Settings



Job settings

Job name

richard-de-prep

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

IAM role

Amazon SageMaker requires permissions to call other services on your behalf. Choose a role or let us create a role that has the **AmazonSageMakerFullAccess** IAM policy attached.

AmazonSageMaker-ExecutionRole-20230101T194235 ▼

Create role using the role creation wizard

Algorithm options

Use an Amazon SageMaker built-in algorithm, your own algorithm, or a third-party algorithm from AWS Marketplace.

▼ **Algorithm source**

● Amazon SageMaker built-in algorithm Learn more

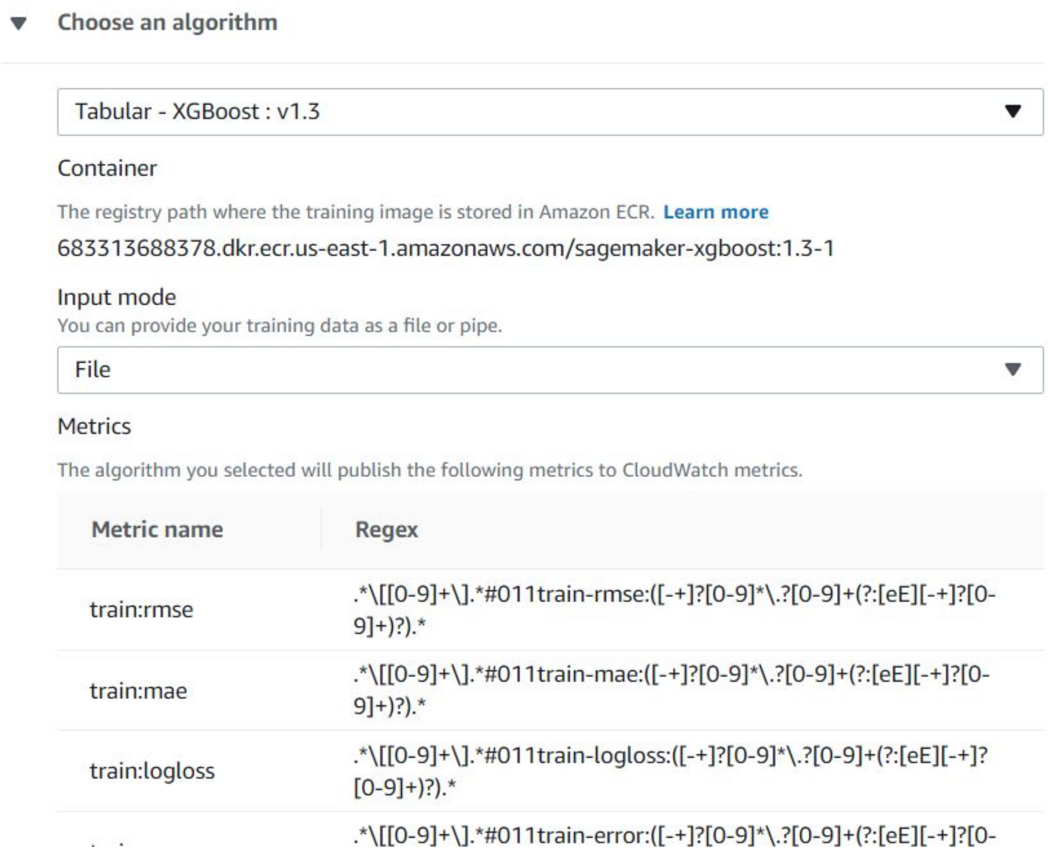
○ Your own algorithm resource

○ Your own algorithm container in ECR Learn more

○ An algorithm subscription from AWS Marketplace

▼ **Choose an algorithm**

Train Job Settings



▼ Choose an algorithm

Tabular - XGBoost : v1.3

Container

The registry path where the training image is stored in Amazon ECR. Learn more

683313688378.dkr.ecr.us-east-1.amazonaws.com/sagemaker-xgboost:1.3-1

Input mode

You can provide your training data as a file or pipe.

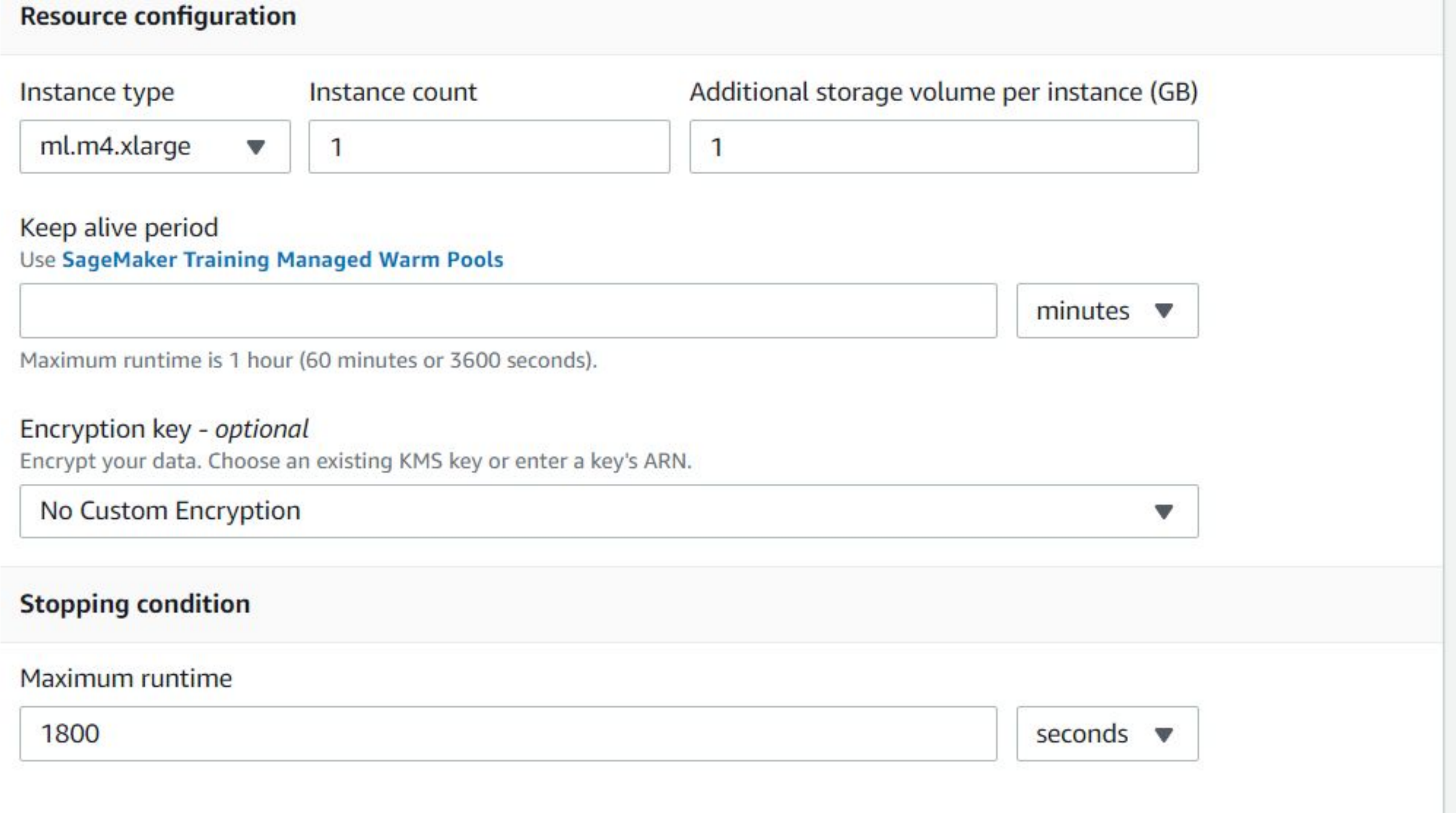
File

Metrics

The algorithm you selected will publish the following metrics to CloudWatch metrics.

Metric name	Regex
train:rmse	.*\[[0-9]+\].*#011train-rmse:([-+]?[0-9]*\.?[0-9]+(?:[eE][-+]?[0-9]+)?).*
train:mae	.*\[[0-9]+\].*#011train-mae:([-+]?[0-9]*\.?[0-9]+(?:[eE][-+]?[0-9]+)?).*
train:logloss	.*\[[0-9]+\].*#011train-logloss:([-+]?[0-9]*\.?[0-9]+(?:[eE][-+]?[0-9]+)?).*
	.*\[[0-9]+\].*#011train-error:([-+]?[0-9]*\.?[0-9]+(?:[eE][-+]?[0-

Train Job Settings



Resource configuration

Instance type: ml.m4.xlarge ▼

Instance count: 1

Additional storage volume per instance (GB): 1

Keep alive period

Use **SageMaker Training Managed Warm Pools**

minutes ▼

Maximum runtime is 1 hour (60 minutes or 3600 seconds).

Encryption key - *optional*

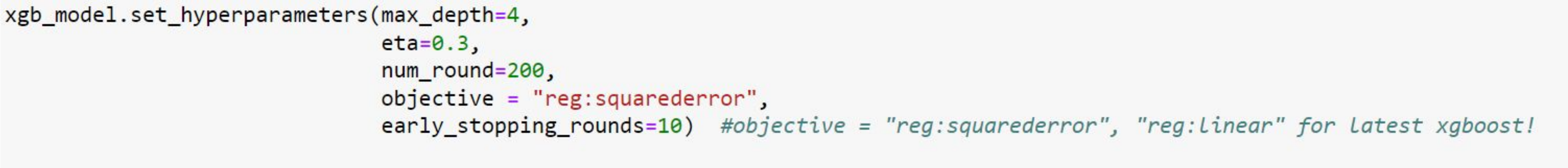
Encrypt your data. Choose an existing KMS key or enter a key's ARN.

No Custom Encryption ▼

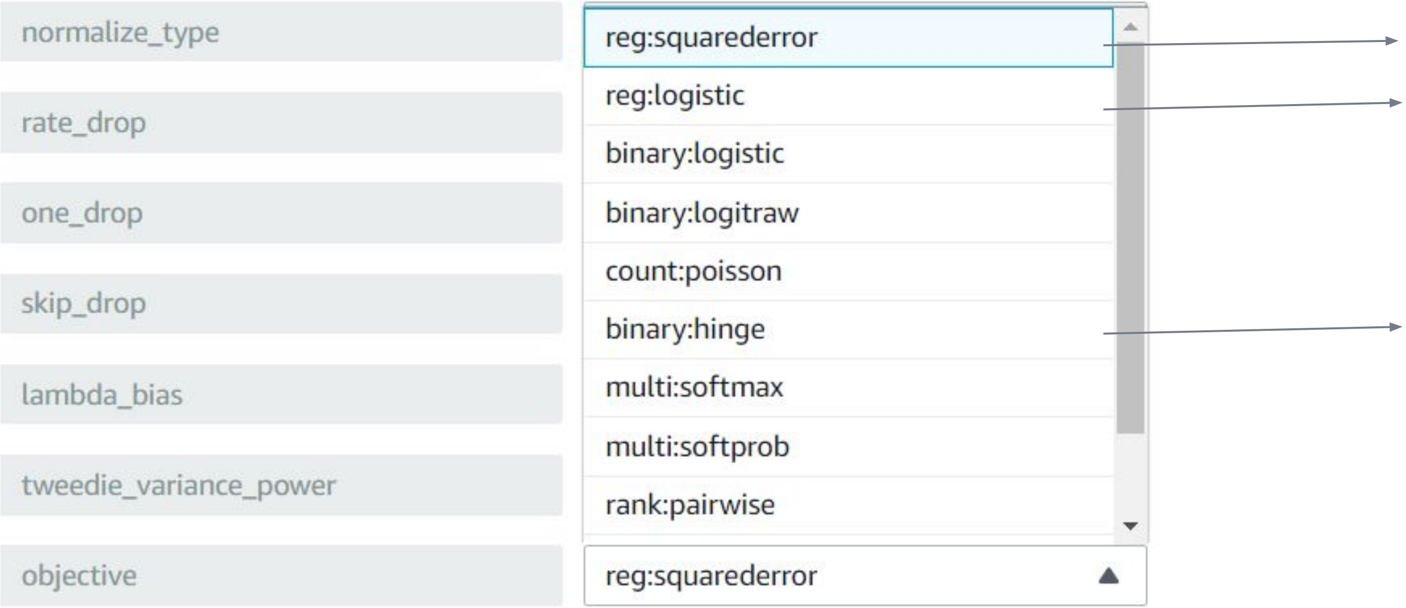
Stopping condition

Maximum runtime: 1800 seconds ▼

Train Job Settings



```
xgb_model.set_hyperparameters(max_depth=4,
                              eta=0.3,
                              num_round=200,
                              objective = "reg:squarederror",
                              early_stopping_rounds=10)  #objective = "reg:squarederror", "reg:linear" for latest xgboost!
```



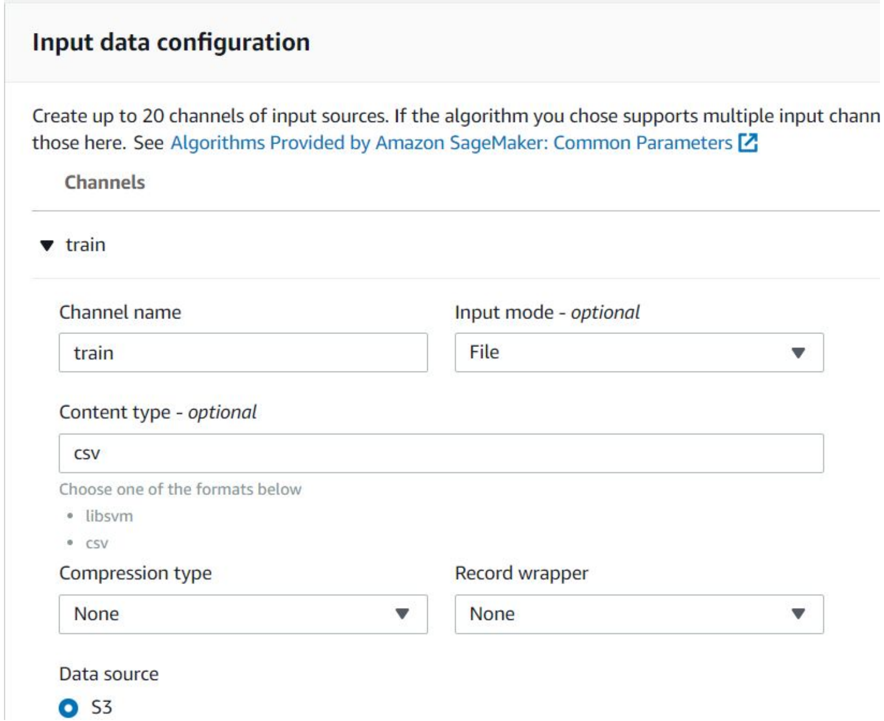
normalize_type	reg:squarederror
rate_drop	reg:logistic
one_drop	binary:logistic
skip_drop	binary:logitraw
lambda_bias	count:poisson
tweedie_variance_power	binary:hinge
	multi:softmax
	multi:softprob
	rank:pairwise
objective	reg:squarederror ▲

regression problems

classification problems

binary classification

Train Job Settings



Input data configuration

Create up to 20 channels of input sources. If the algorithm you chose supports multiple input chann
those here. See Algorithms Provided by Amazon SageMaker: Common Parameters

Channels

▼ train

Channel name: train

Input mode - *optional*: File ▼

Content type - *optional*: csv

Choose one of the formats below

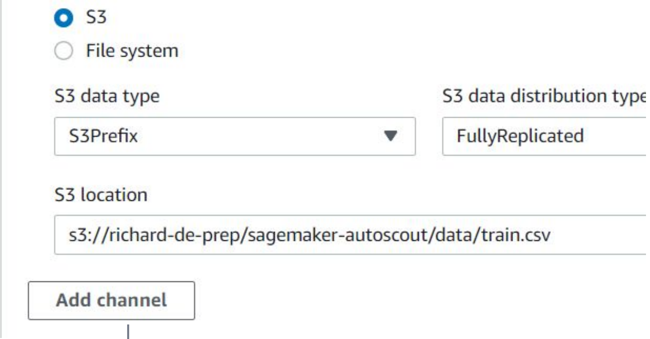
- libsvm
- csv

Compression type: None ▼

Record wrapper: None ▼

Data source

● S3



● S3

○ File system

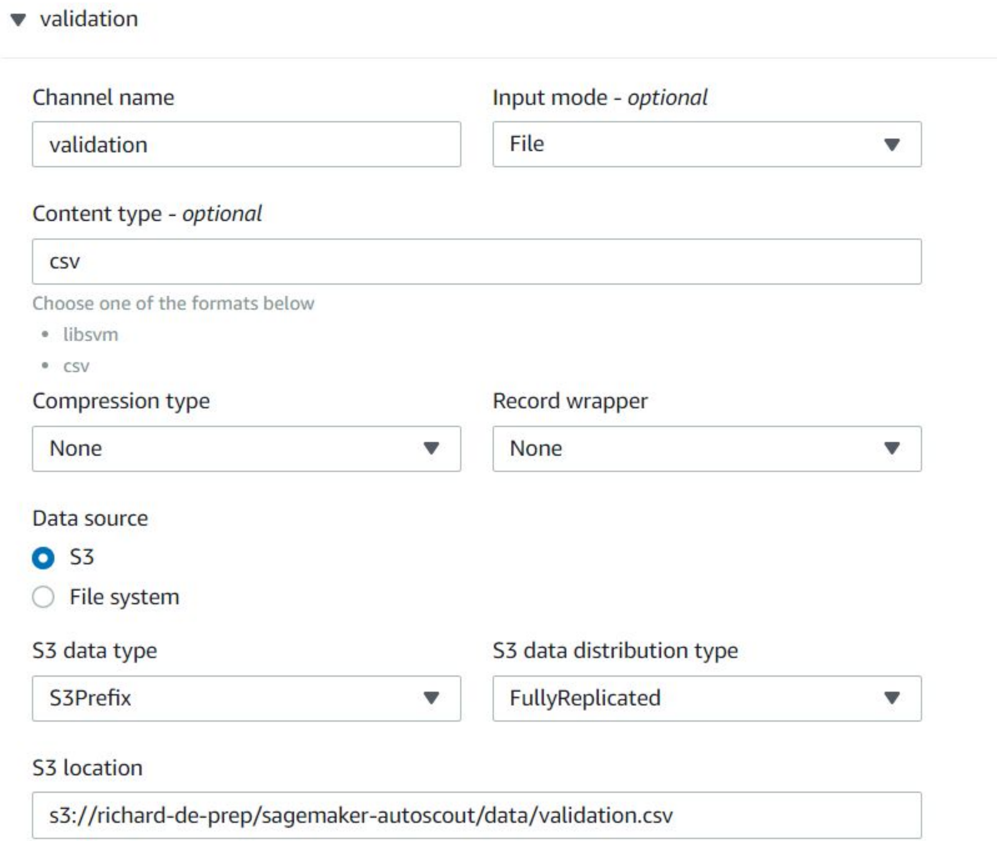
S3 data type: S3Prefix ▼

S3 data distribution type: FullyReplicated

S3 location: s3://richard-de-prep/sagemaker-autoscout/data/train.csv

Don’t forget to add channel for validation data

Train Job Settings



▼ validation

Channel name

validation

Input mode - *optional*

File ▼

Content type - *optional*

csv

Choose one of the formats below

- libsvm
- csv

Compression type

None ▼

Record wrapper

None ▼

Data source

● S3

○ File system

S3 data type

S3Prefix ▼

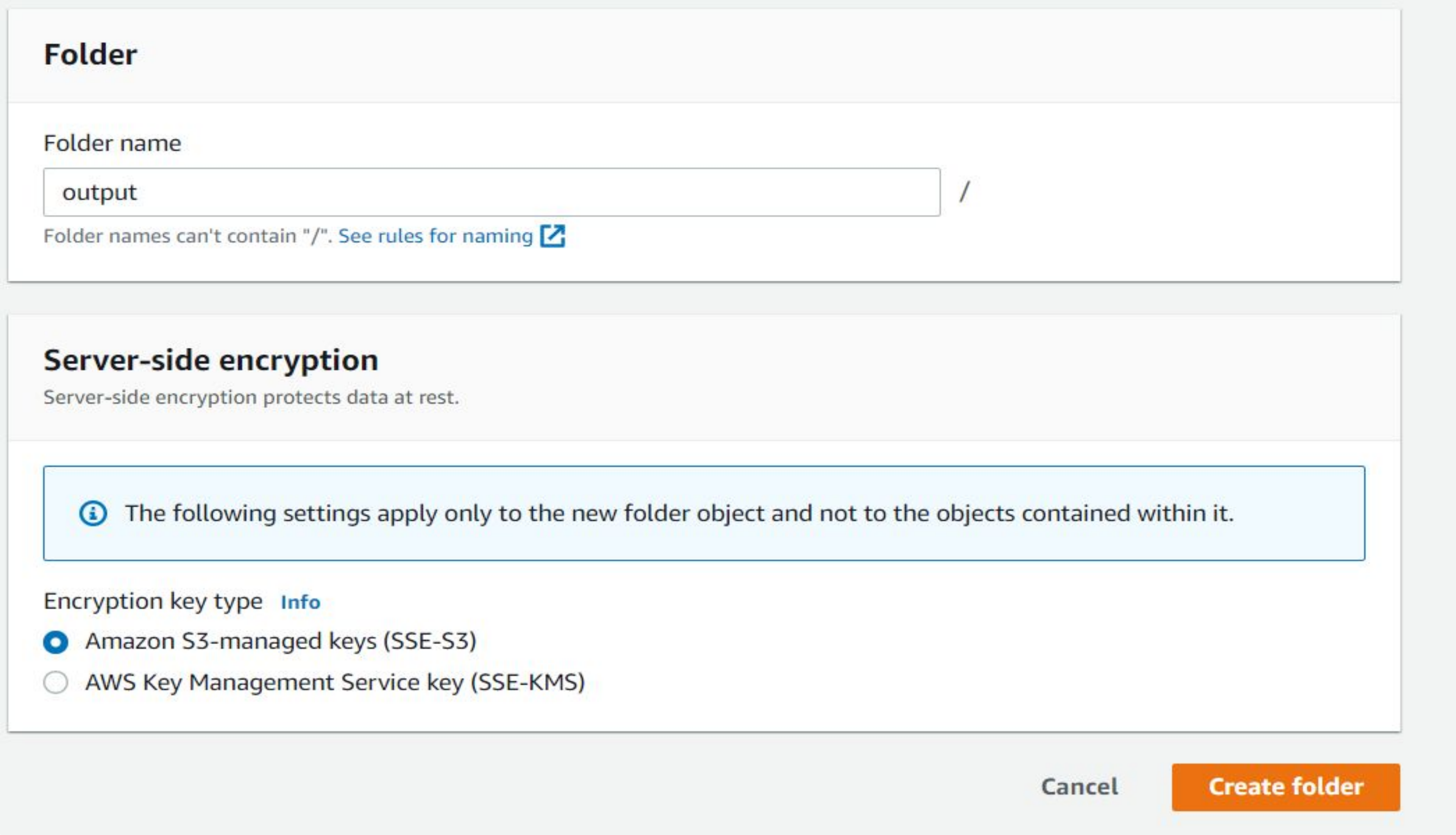
S3 data distribution type

FullyReplicated ▼

S3 location

s3://richard-de-prep/sagemaker-autoscout/data/validation.csv

Create Output Folder from S3



Folder

Folder name

output /

Folder names can't contain "/". See rules for naming

Server-side encryption

Server-side encryption protects data at rest.

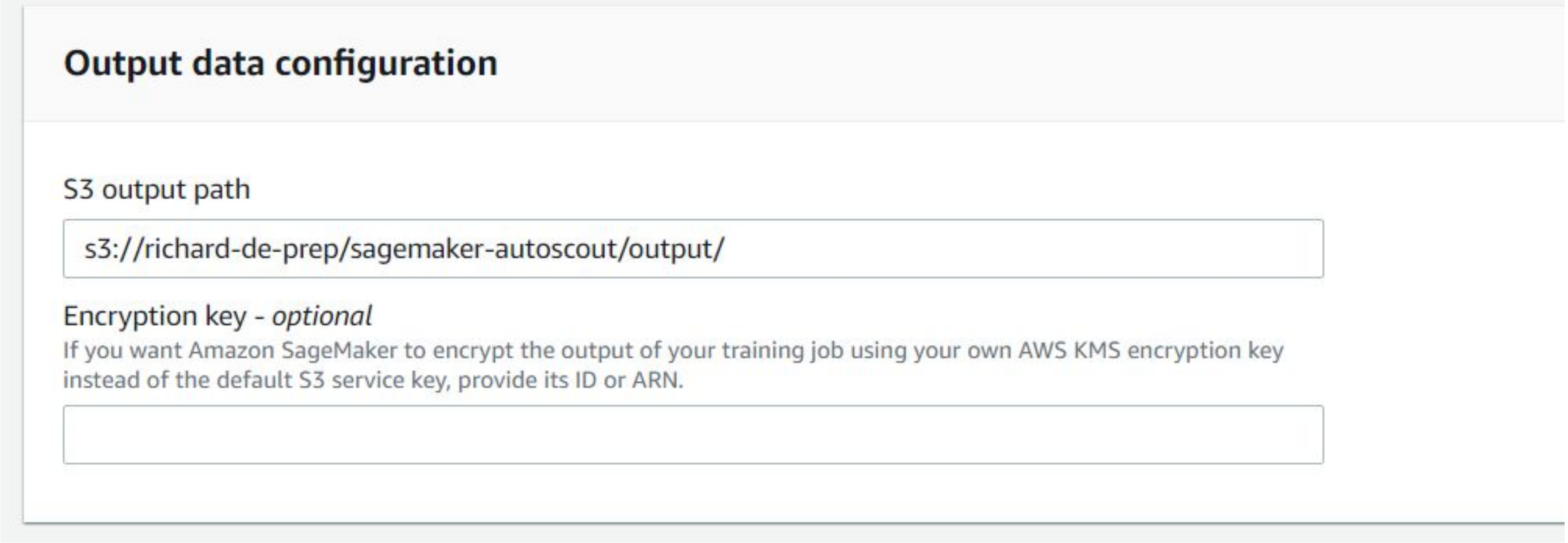
The following settings apply only to the new folder object and not to the objects contained within it.

Encryption key type Info

● Amazon S3-managed keys (SSE-S3)

○ AWS Key Management Service key (SSE-KMS)

Train Job Settings



Output data configuration

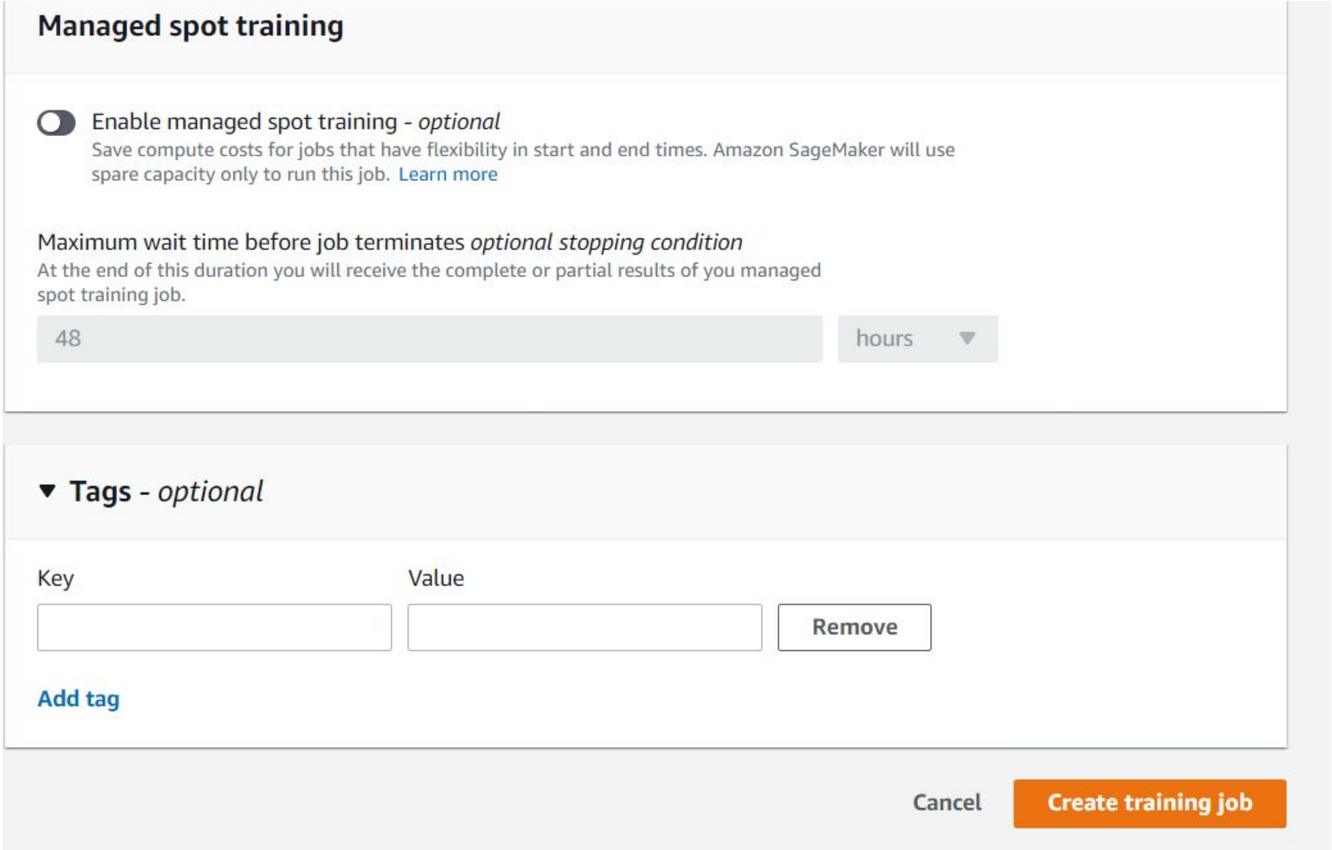
S3 output path

s3://richard-de-prep/sagemaker-autoscout/output/

Encryption key - *optional*

If you want Amazon SageMaker to encrypt the output of your training job using your own AWS KMS encryption key instead of the default S3 service key, provide its ID or ARN.

Train Job Settings



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.

48 hours ▼

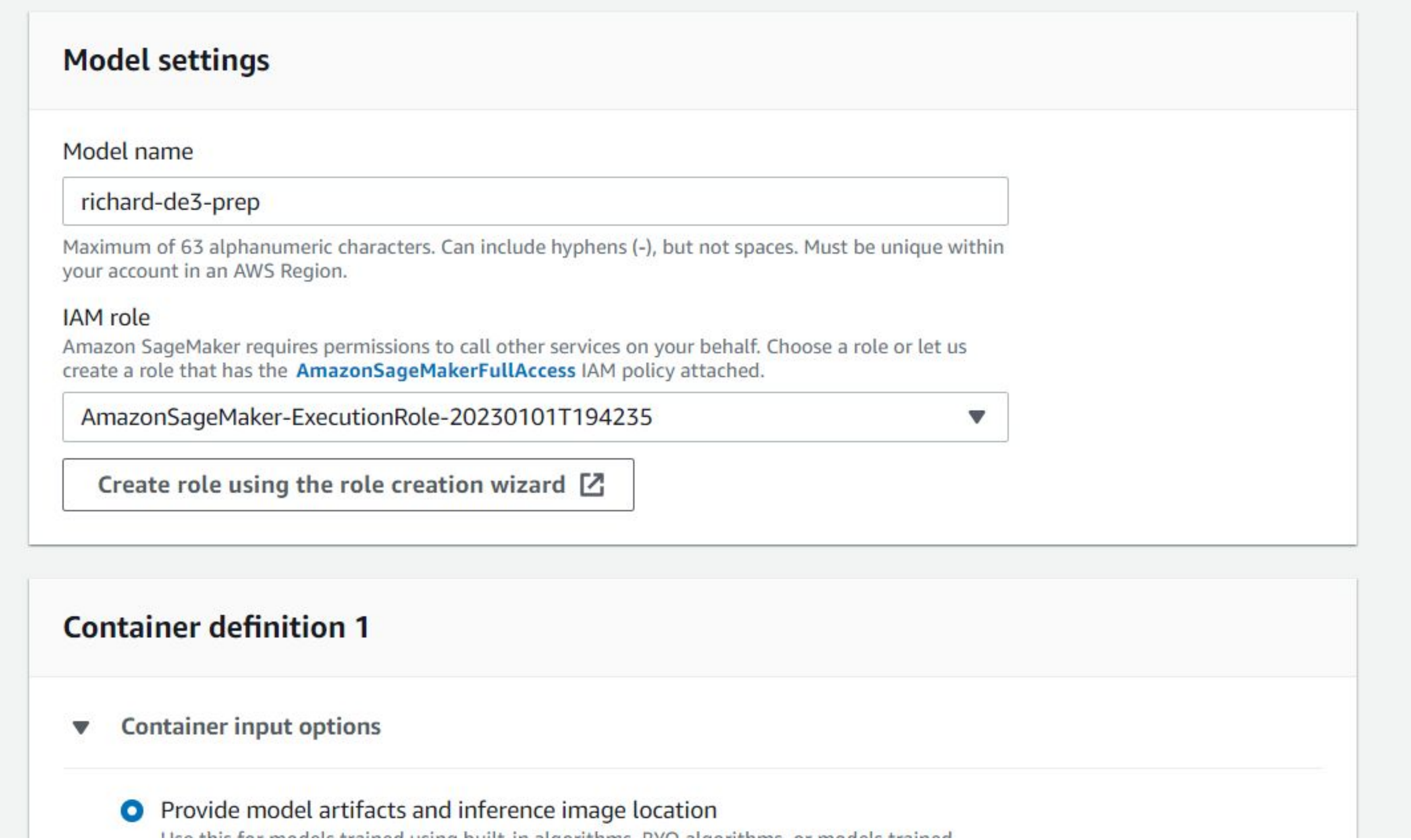
▼ Tags - *optional*

Key Value Remove

Add tag

Cancel Create training job

Model Settings



Model settings

Model name

richard-de3-prep

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

IAM role

Amazon SageMaker requires permissions to call other services on your behalf. Choose a role or let us create a role that has the **AmazonSageMakerFullAccess** IAM policy attached.

AmazonSageMaker-ExecutionRole-20230101T194235 ▼

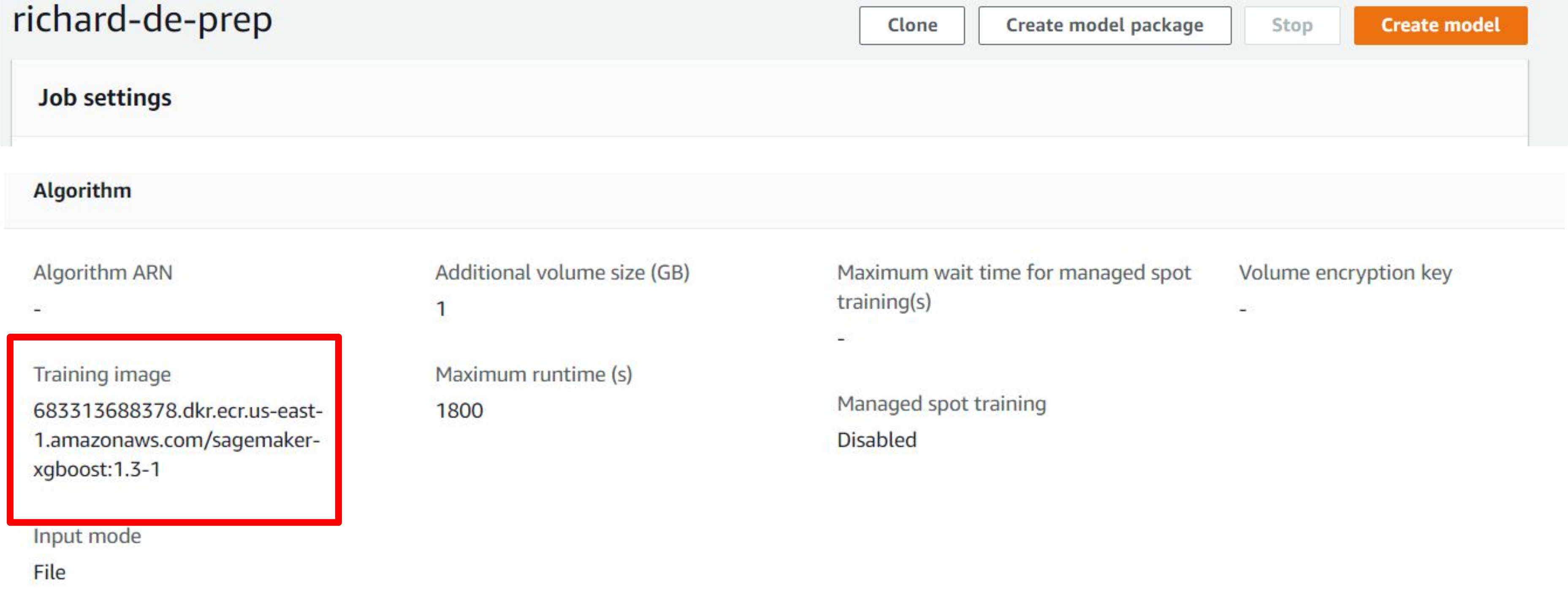
Container definition 1

▼ **Container input options**

● Provide model artifacts and inference image location

Use this for models trained using built-in algorithms, BYO algorithms, or models trained

Model Settings



richard-de-prep

Job settings

Algorithm

Algorithm ARN
-

Additional volume size (GB)
1

Maximum wait time for managed spot training(s)
-

Volume encryption key
-

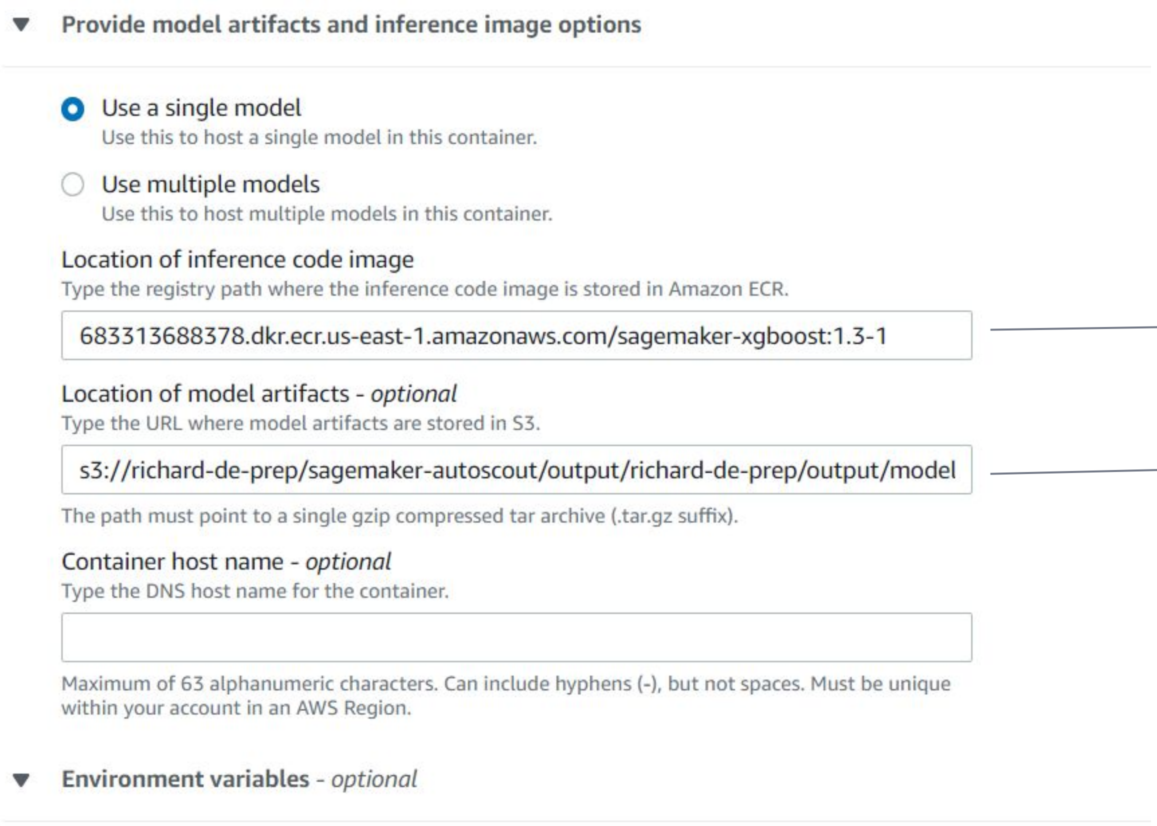
Training image
683313688378.dkr.ecr.us-east-1.amazonaws.com/sagemaker-xgboost:1.3-1

Maximum runtime (s)
1800

Managed spot training
Disabled

Input mode
File

Model Settings



▼ **Provide model artifacts and inference image options**

● Use a single model
Use this to host a single model in this container.

○ Use multiple models
Use this to host multiple models in this container.

Location of inference code image
Type the registry path where the inference code image is stored in Amazon ECR.

683313688378.dkr.ecr.us-east-1.amazonaws.com/sagemaker-xgboost:1.3-1

Location of model artifacts - *optional*
Type the URL where model artifacts are stored in S3.

s3://richard-de-prep/sagemaker-autoscout/output/richard-de-prep/output/model

The path must point to a single gzip compressed tar archive (.tar.gz suffix).

Container host name - *optional*
Type the DNS host name for the container.

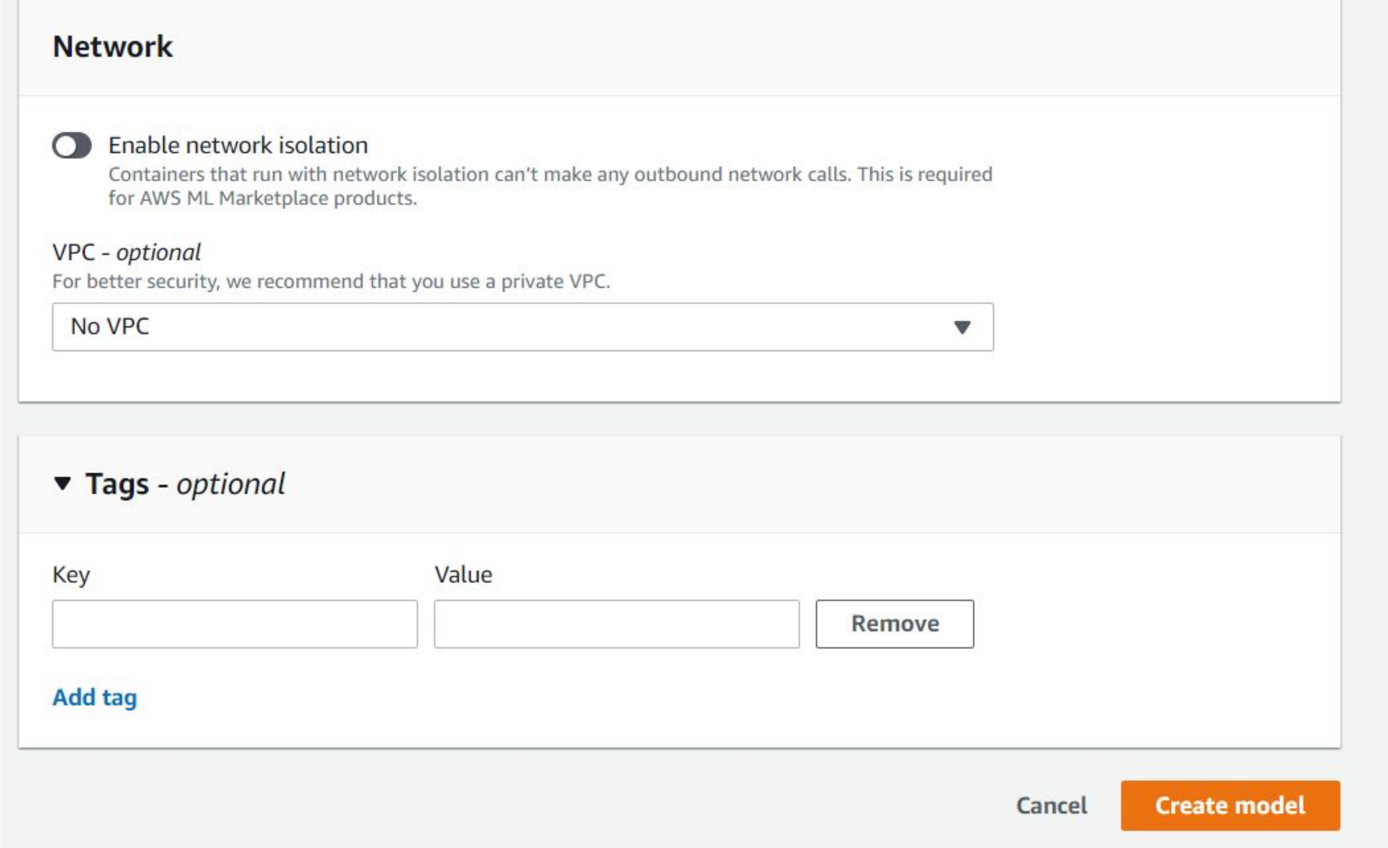
Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

▼ **Environment variables** - *optional*

from training job

from s3 (modeltargz uri)

Model Settings



Network

Enable network isolation
Containers that run with network isolation can't make any outbound network calls. This is required for AWS ML Marketplace products.

VPC - *optional*
For better security, we recommend that you use a private VPC.

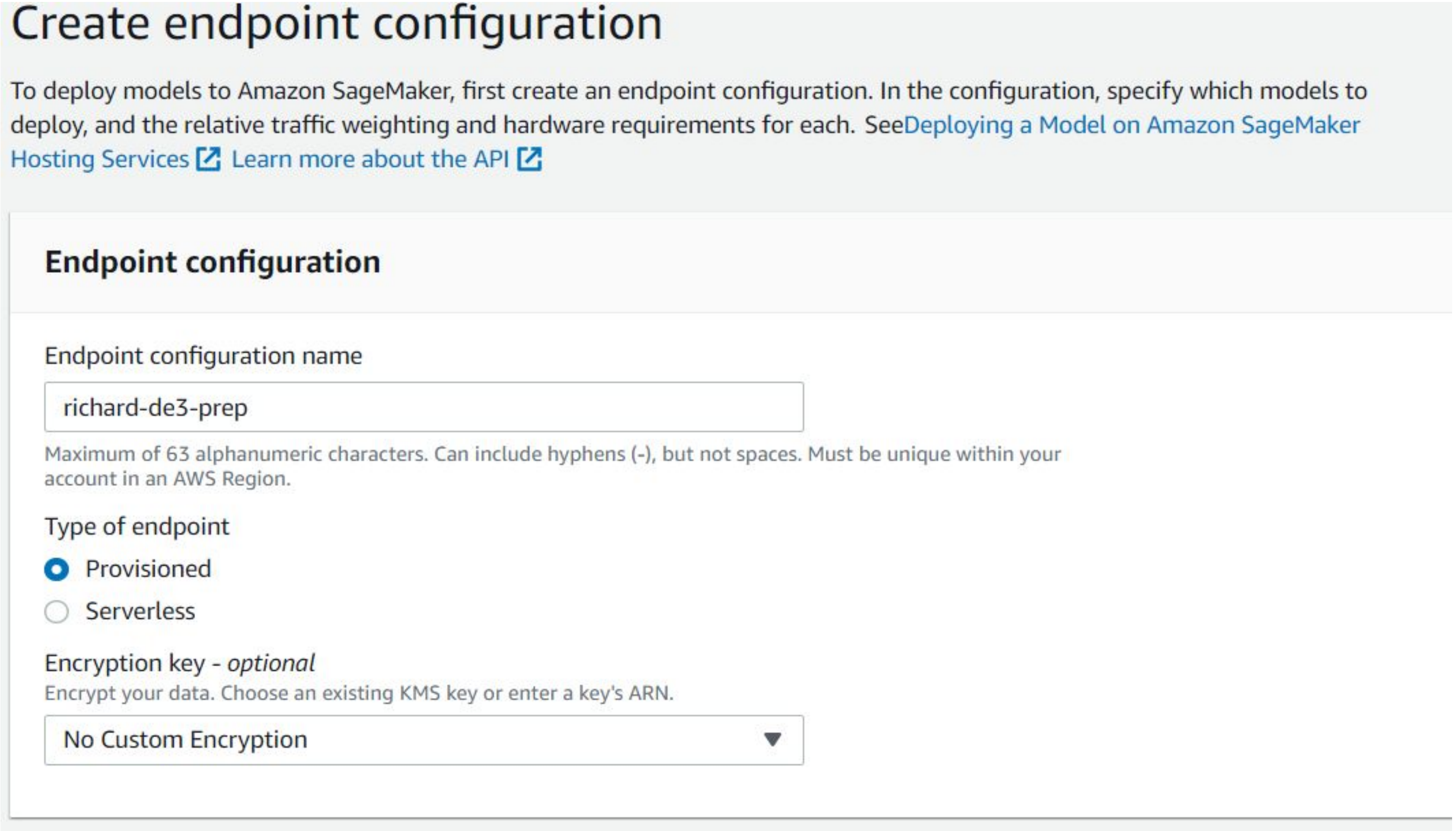
No VPC

▼ Tags - *optional*

Key

Value

Endpoint Config Settings



Create endpoint configuration

To deploy models to Amazon SageMaker, first create an endpoint configuration. In the configuration, specify which models to deploy, and the relative traffic weighting and hardware requirements for each. SeeDeploying a Model on Amazon SageMaker Hosting Services ↗ Learn more about the API ↗

Endpoint configuration

Endpoint configuration name

richard-de3-prep

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

Type of endpoint

● Provisioned

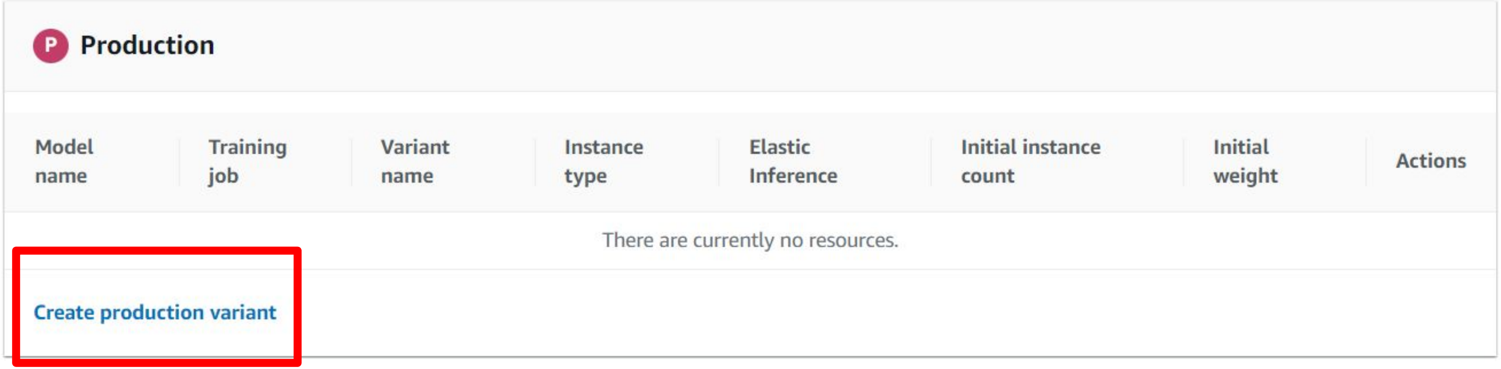
○ Serverless

Encryption key - *optional*

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

No Custom Encryption ▼

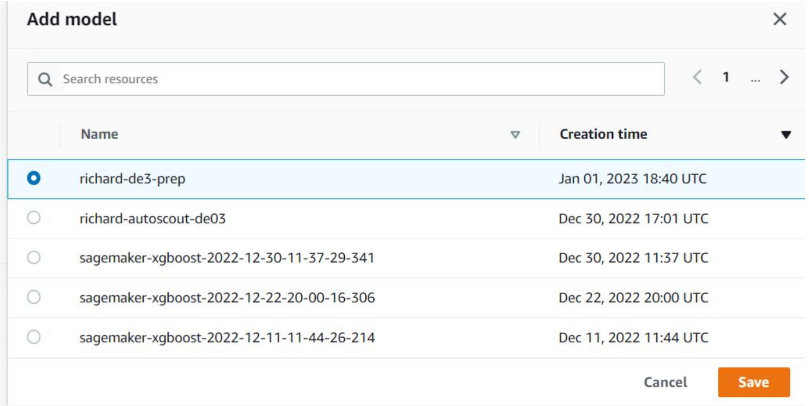
Endpoint Config Settings



P **Production**

Model name	Training job	Variant name	Instance type	Elastic Inference	Initial instance count	Initial weight	Actions
There are currently no resources.							

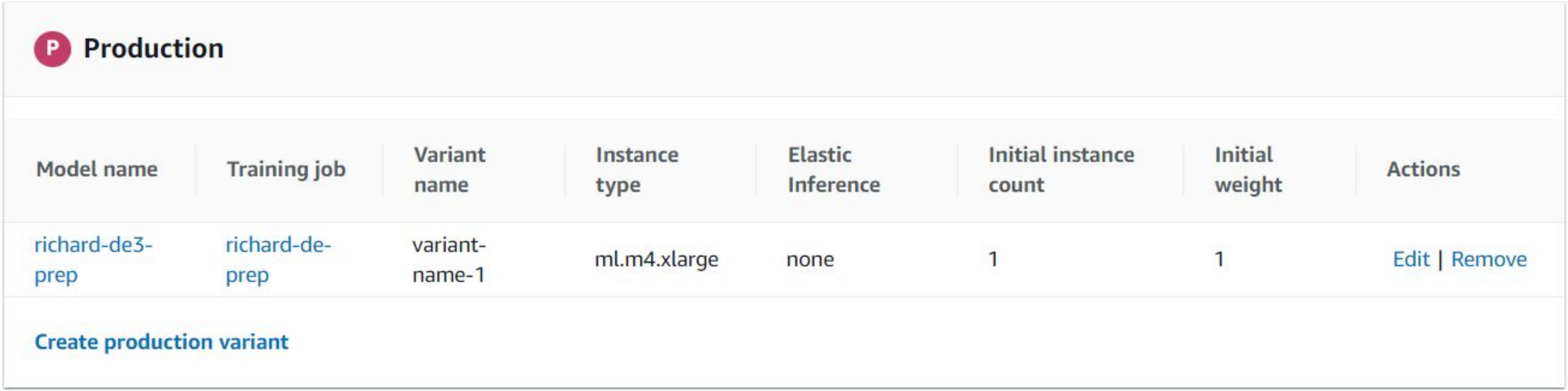
Create production variant



Add model

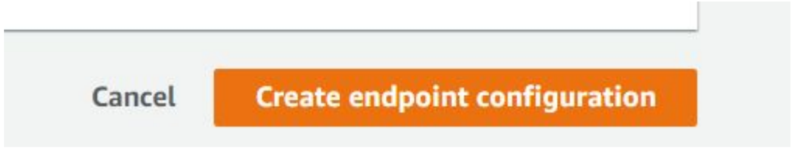
	Name	Creation time
●	richard-de3-prep	Jan 01, 2023 18:40 UTC
○	richard-autoscout-de03	Dec 30, 2022 17:01 UTC
○	sagemaker-xgboost-2022-12-30-11-37-29-341	Dec 30, 2022 11:37 UTC
○	sagemaker-xgboost-2022-12-22-20-00-16-306	Dec 22, 2022 20:00 UTC
○	sagemaker-xgboost-2022-12-11-11-44-26-214	Dec 11, 2022 11:44 UTC

Endpoint Config Settings

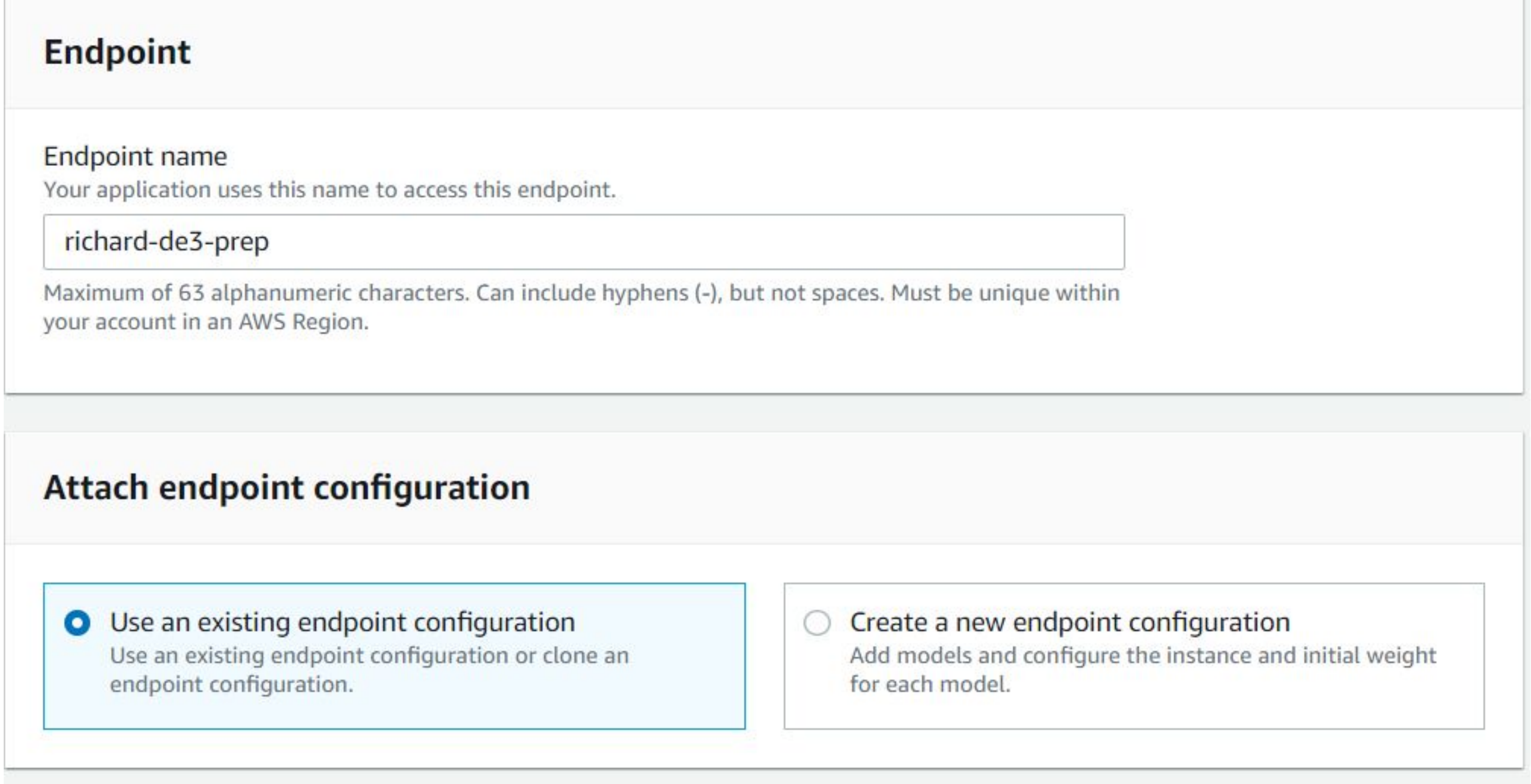


Production

Model name	Training job	Variant name	Instance type	Elastic Inference	Initial instance count	Initial weight	Actions
richard-de3-prep	richard-de-prep	variant-name-1	ml.m4.xlarge	none	1	1	Edit | Remove



Endpoint Settings



Endpoint

Endpoint name

Your application uses this name to access this endpoint.

richard-de3-prep

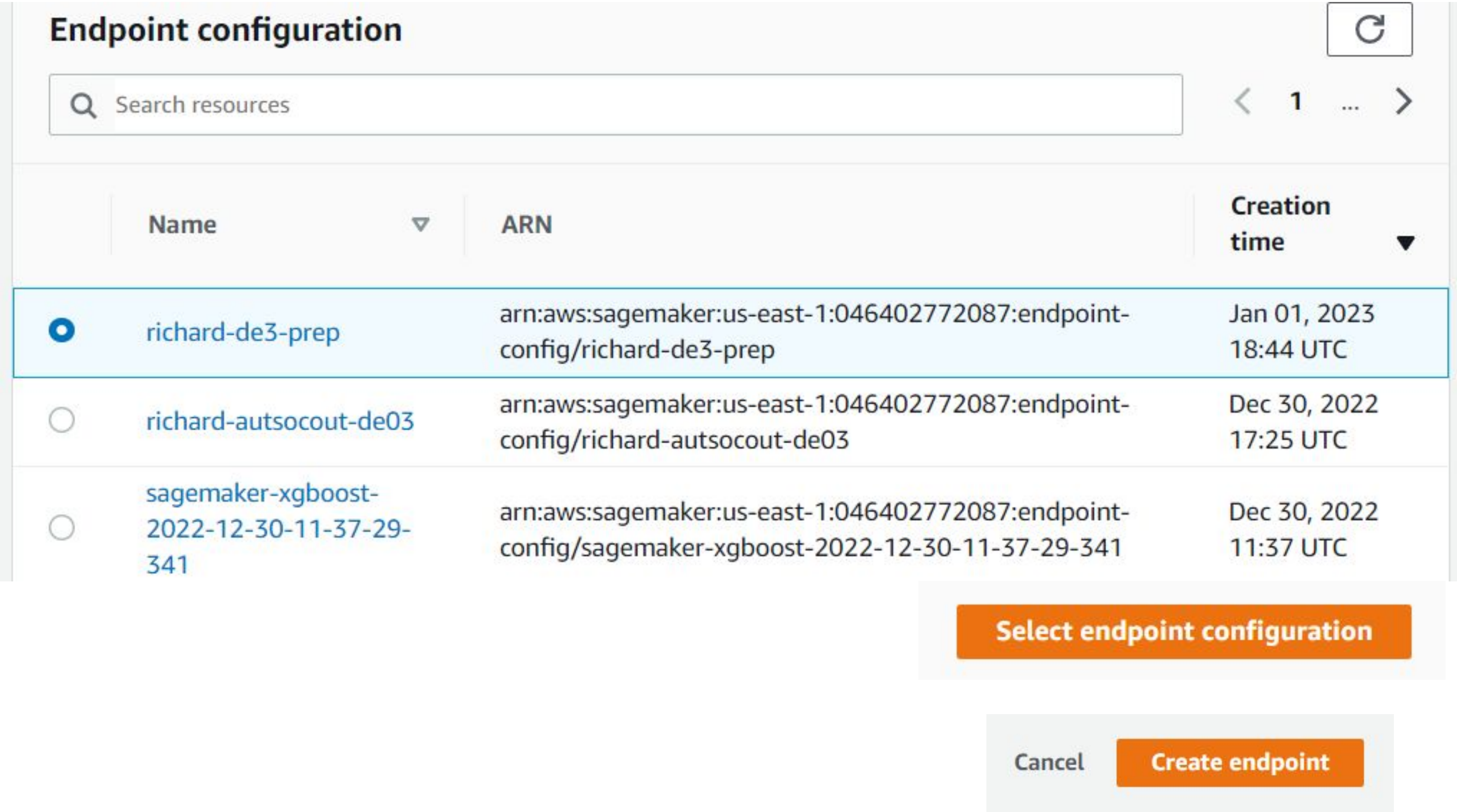
Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Attach endpoint configuration

● Use an existing endpoint configuration
Use an existing endpoint configuration or clone an endpoint configuration.

○ Create a new endpoint configuration
Add models and configure the instance and initial weight for each model.

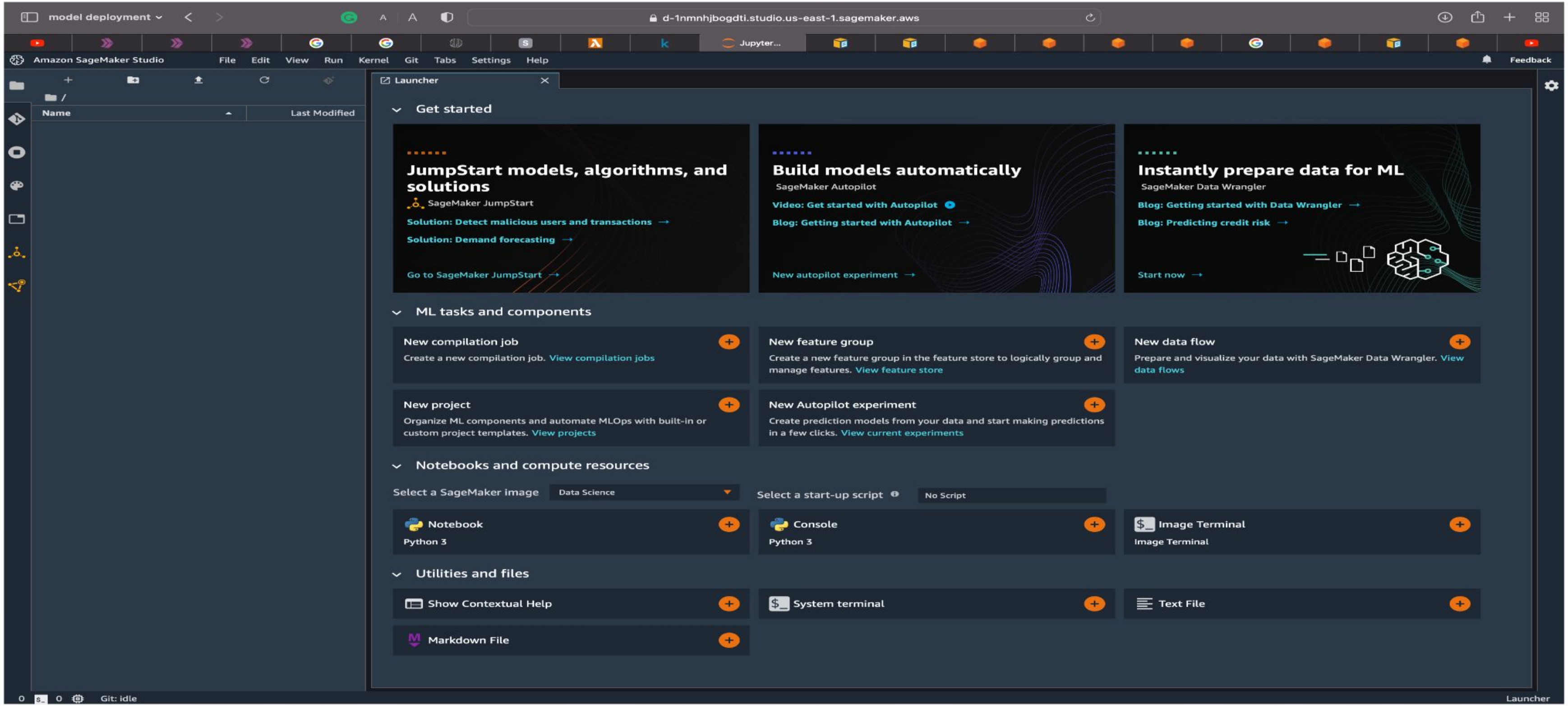
Endpoint Settings



Endpoint configuration

	Name	ARN	Creation time
●	richard-de3-prep	arn:aws:sagemaker:us-east-1:046402772087:endpoint-config/richard-de3-prep	Jan 01, 2023 18:44 UTC
○	richard-autsocout-de03	arn:aws:sagemaker:us-east-1:046402772087:endpoint-config/richard-autsocout-de03	Dec 30, 2022 17:25 UTC
○	sagemaker-xgboost-2022-12-30-11-37-29-341	arn:aws:sagemaker:us-east-1:046402772087:endpoint-config/sagemaker-xgboost-2022-12-30-11-37-29-341	Dec 30, 2022 11:37 UTC

SageMaker Studio



Get started

JumpStart models, algorithms, and solutions
SageMaker JumpStart
Solution: Detect malicious users and transactions →
Solution: Demand forecasting →
Go to SageMaker JumpStart →

Build models automatically
SageMaker Autopilot
Video: Get started with Autopilot
Blog: Getting started with Autopilot →
New autopilot experiment →

Instantly prepare data for ML
SageMaker Data Wrangler
Blog: Getting started with Data Wrangler →
Blog: Predicting credit risk →
Start now →

ML tasks and components

New compilation job
Create a new compilation job. View compilation jobs

New feature group
Create a new feature group in the feature store to logically group and manage features. View feature store

New data flow
Prepare and visualize your data with SageMaker Data Wrangler. View data flows

New project
Organize ML components and automate MLOps with built-in or custom project templates. View projects

New Autopilot experiment
Create prediction models from your data and start making predictions in a few clicks. View current experiments

Notebooks and compute resources

Select a SageMaker image: Data Science

Select a start-up script: No Script

Notebook
Python 3

Console
Python 3

Image Terminal
Image Terminal

Utilities and files

Show Contextual Help

System terminal

Text File

Markdown File