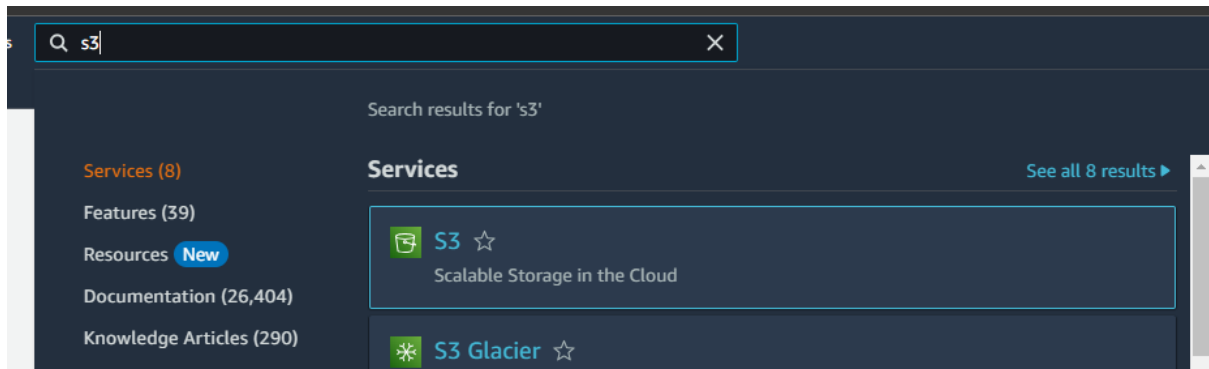
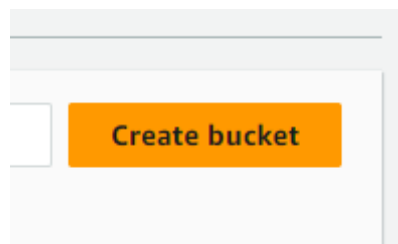


Practical 10: Create a Labeling Job

Step 1: Open S3



Step 2: Create a bucket



Create bucket [Info](#)

Buckets are containers for data stored in S3.

General configuration

AWS Region
Asia Pacific (Mumbai) ap-south-1

Bucket name [Info](#)

Bucket name must be unique within the global namespace and follow the bucket naming rules. [See rules for bucket naming](#)

Copy settings from existing bucket - *optional*
Only the bucket settings in the following configuration are copied.

Format: s3://bucket/prefix

✔ **Successfully created bucket "ground-truth-practical"**
To upload files and folders, or to configure additional bucket settings, choose **View details**.

Step 3: Save two images at publicly available HTTP URLs. The images are used when creating instructions for completing a labeling task. The images should have an aspect ratio of around 2:1. For this exercise, the content of the images is not important.

[Amazon S3](#) > [Buckets](#) > [ground-truth-practical](#) > Upload

Upload Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.

Files and folders (2 Total, 627.4 KB)
All files and folders in this table will be uploaded.

[Remove](#) [Add files](#) [Add folder](#)

☐

Name

▼

Folder

☐

cinematic.jpg

-

☐

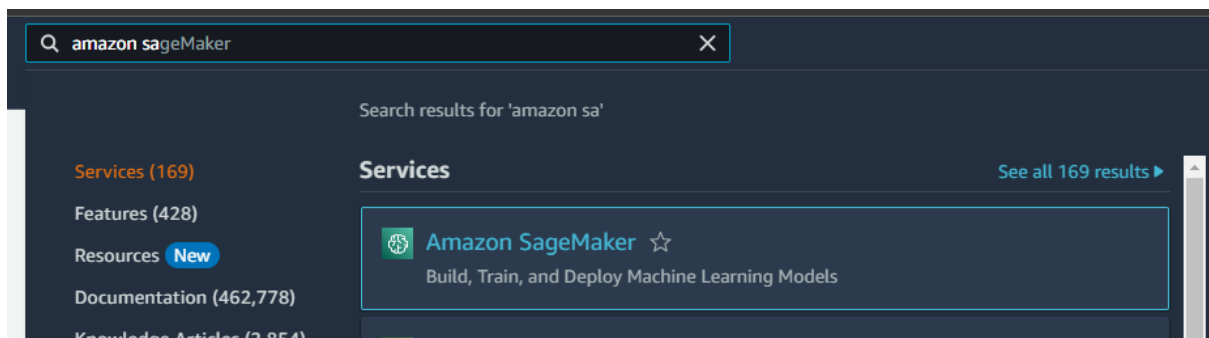
house-of-cards-aspect-ratio.jpg

-

Step 4: Successfully updated

Files and folders (2 Total, 627.4 KB)						
<input type="text" value="Find by name"/>						
Name	Folder	▼	Type	▼	Size	▼
cinematic.jpg	-		image/jpeg		240.6 KB	
house-of-ca...	-		image/jpeg		386.8 KB	

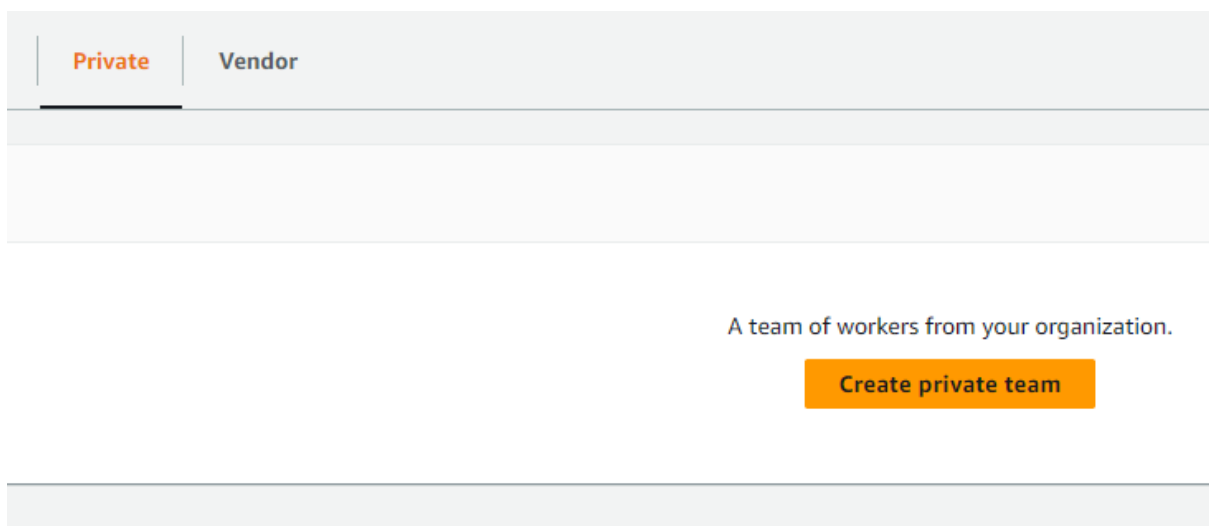
Step 5: Open the SageMaker console



Step 6: From the left navigation, choose **Labeling workforces**

- ▼ Ground Truth
 - Labeling jobs
 - Labeling datasets
 - Labeling workforces**
- Notebook

Step 7: create a private team



Step 8: Enter a name and description for your team. Add email , organization name

[Amazon SageMaker](#) > [Labeling workforces](#) > Create private team

Create private team

Private team creation

☒ **Create a private team with AWS Cognito**
Create a private work team by sending email invitations to new workers or importing workers from existing Amazon Cognito user groups.

☐ **Create a private team with OpenID Connect (OIDC)**
Create a private work team with your own identity provider (IdP). Your IdP must support OIDC user groups.

Team details

Team name
Give your work team a descriptive name. This name can't be changed later.

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

Add workers [Info](#)

Add workers to your private work team by adding worker email addresses or importing workers from existing Amazon Cognito user groups.

☒ **Invite new workers by email**

☐ **Import workers from existing Amazon Cognito user groups**

Cancel

Create private team

✓

Private team 'Msc-work-team' was successfully created.

[Amazon SageMaker](#) > Labeling workforces

Step 9: Add workers

Add workers [Info](#)
Add workers to your private work team by adding worker email addresses or importing workers from existing Amazon Cognito user groups.

☒ Invite new workers by email

☐ Import workers from existing Amazon Cognito user groups

Email addresses
We send an invitation with instructions to each of the worker email addresses that you add here.

Use a comma between addresses. You can add up to 50 workers.

Organization name
We use this information to customize the email that we send to the workers.

Contact email
Workers use this address to report issues related to the task.

We send an email with the login details to all the workers added to your team.

Email Invitation
We send an email with the login details to all the workers added to your team.

Preview invitation

Step 10: Open the SageMaker console

aws

Services

Search

[Alt+S]

EC2

Amazon SageMaker

Getting started

Studio

Studio Lab

Canvas

RStudio

Admin configurations

Domains

Role manager

Images

Lifecycle configurations

SageMaker dashboard

Search

Amazon SageMaker > Domains

Domains

Info

In SageMaker, a domain is an environment for your team to access SageMaker resources. A domain can contain either one or multiple domains.

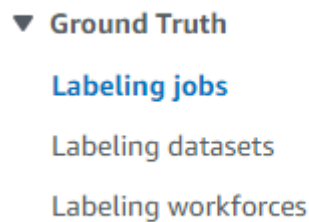
Domains (0)

Info

Find domain name

Name	Id	Status
------	----	--------

Step 11: From the left navigation, choose **Labeling jobs**



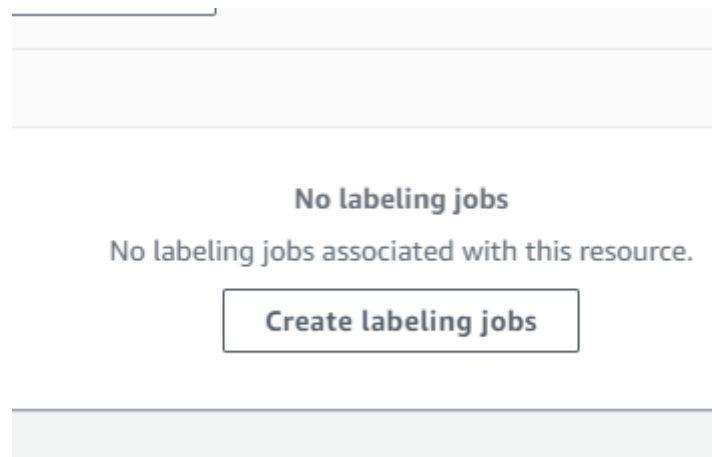
▼ Ground Truth

Labeling jobs

Labeling datasets

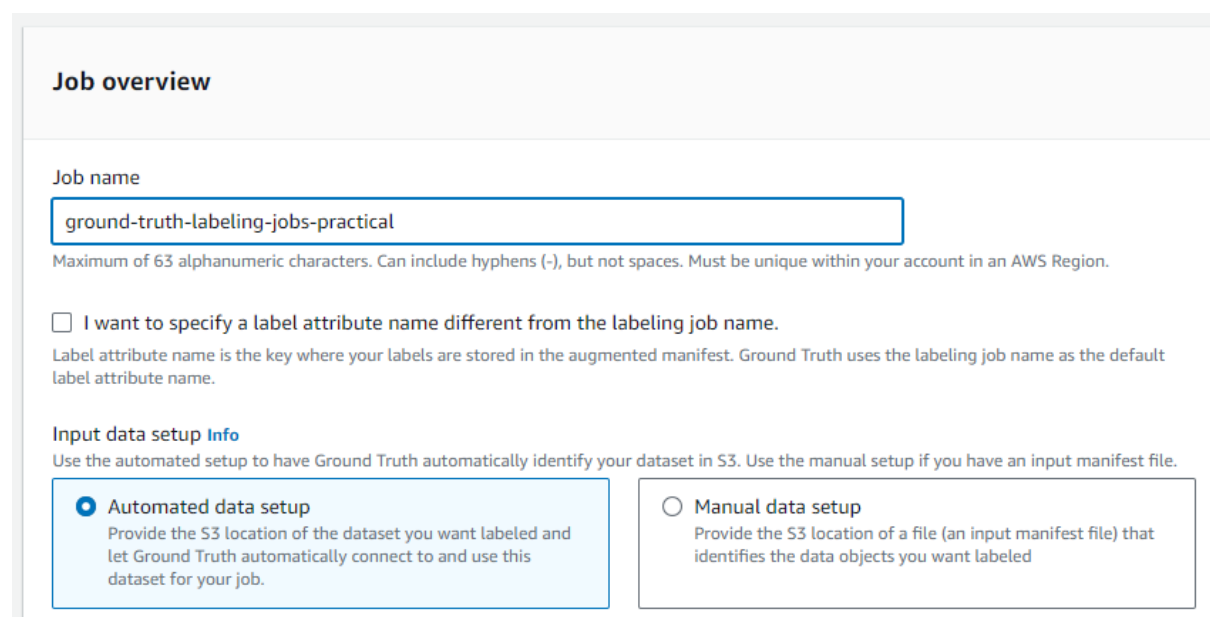
Labeling workforces

Step 12: Choose **Create labeling job** to start the job creation process



Step 13: **Job name** – Give the labeling job a name that describes the job. This name is shown in your job list. The name must be unique in your account in an AWS Region.

Input data setup – Select Automated data setup. This option allows you to automatically connect to your input data in S3.



Job overview

Job name

ground-truth-labeling-jobs-practical

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

☐ I want to specify a label attribute name different from the labeling job name.

Label attribute name is the key where your labels are stored in the augmented manifest. Ground Truth uses the labeling job name as the default label attribute name.

Input data setup [Info](#)

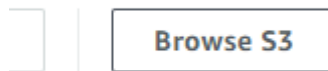
Use the automated setup to have Ground Truth automatically identify your dataset in S3. Use the manual setup if you have an input manifest file.

☒ **Automated data setup**
Provide the S3 location of the dataset you want labeled and let Ground Truth automatically connect to and use this dataset for your job.

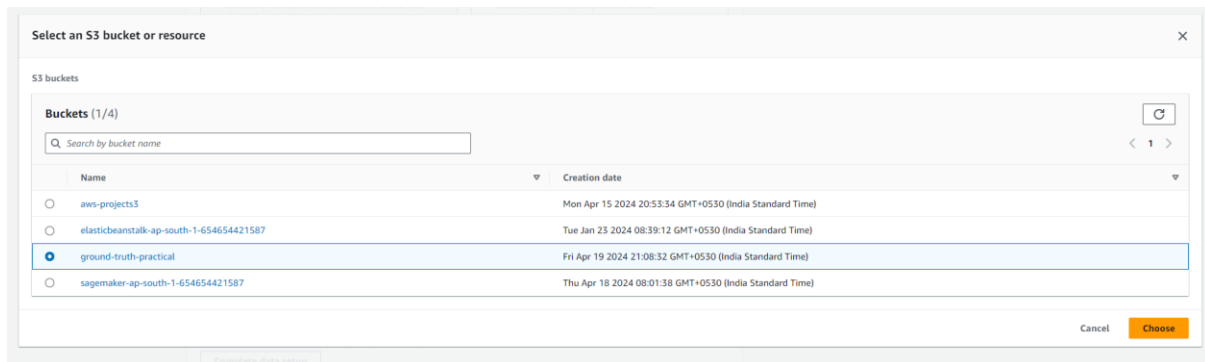
☐ **Manual data setup**
Provide the S3 location of a file (an input manifest file) that identifies the data objects you want labeled

Step 14: Browse any S3 bucket

will use all data objects in this location for



Step 15: Locate the bucket



Step 16: S3 location for input datasets – Enter the S3 location where you added the images in step 1.

Data type – Use the drop down menu to **select Image**. Ground Truth will use all images found in the S3 location for input datasets as input for your labeling job.

S3 location for output datasets [Info](#)

This is the location in S3 where your labeling job output data is stored.

- ☒ Same location as input dataset
- ☐ Specify a new location

Data type

Image

Supported formats are .jpg, .jpeg, and .png.

▼

Step 17: **IAM role** – Create or choose an IAM role with the AmazonSageMakerFullAccess IAM policy attached.

IAM Role Info
Provide the ID or ARN for your own AWS KMS encryption key for Amazon SageMaker to access your S3 bucket. Choose role with the [AmazonSageMakerFullAccess](#) IAM policy attached.

Choose an option ▲

☒ Create a new role

☐ Enter a custom IAM role ARN

☐ Use existing role

Step 18: Select any S3 bucket

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
 - ☐ Specific S3 buckets
 - ☒ Any S3 bucket
Allow Ground Truth to have access to any bucket and its content in your account.
 - ☐ None
- ☒ Any S3 bucket with "sagemaker" in the name
- ☒ Any S3 object with "sagemaker" in the name
- ☒ Any S3 object with tag "sagemaker" and value "true" [See Object tagging](#)
- ☒ Any S3 buckets with a Bucket Policy allowing access to SageMaker [See S3 bucket policies](#)

Cancel Create

Step 19: Click Complete data setup

Use this button to process and complete your input data setup.

Complete data setup

► **Additional configuration - optional**
Dataset object selection, encryption

Use this button to process and complete your input data setup.

Complete data setup

✓ Input data connection successful. [View more details](#)

► **Additional configuration - optional**

Dataset object selection, encryption

Step 20: Select Image and Bounding Box

Task type [Info](#)

Task category

Select the type of data being labeled to view available task templates for it or select 'Custom' to create your own.

Image ▼

Task selection

Select the task that a human worker will perform to label objects in your dataset.

☐ Image Classification (Single Label)

Get workers to categorize images into individual classes. [Info](#)



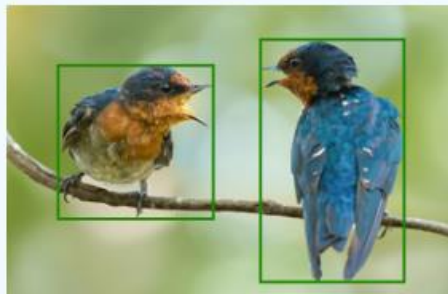
☐ Image Classification (Multi-label)

Get workers to categorize images into one or more classes. [Info](#)



☒ Bounding box

Get workers to draw bounding boxes around specified objects in your images. [Info](#)

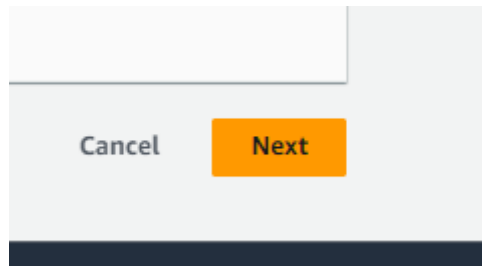


☐ Semantic segmentation

Get workers to draw pixel level labels around specific objects and segments in your images. [Info](#)



Step 21: Click Next



Step 22: In the Workers section, choose Private. Give team name

Select workers and configure tool

All fields are required unless otherwise specified

Workers [Info](#)

Worker types

☐ Amazon Mechanical Turk
An on-demand 24/7 workforce of over 500,000 independent contractors worldwide powered by Amazon Mechanical Turk.

☒ Private
A team of workers that you have sourced yourself, including your own employees or contractors for handling data that needs to stay within your organization.

☐ Vendor managed
A curated list of third party vendors that specialize in providing data labeling services, available via the AWS Marketplace.

Private teams
Choose from the teams you created in the private workforce or if you need to create a new team, save your progress and go to Labeling workforces to create a new one.

Msc-work-team ▼

Task timeout
The maximum time a worker can work in a single task. Please see [here](#) for information on default and maximum values.

1 hours 0 mins 0 secs

Task expiration time
The amount of time that a task remains available to workers before expiring. Please see [here](#) for information on default and maximum values.

10 days 0 hours 0 mins
0 secs

☐ Enable automated data labeling [Info](#)
Amazon SageMaker will automatically label a portion of your dataset. It will train a model in your AWS account using Built-in Algorithm and your dataset. When you enable this, training jobs use new computing resources on your behalf. For cost information, See SageMaker [pricing](#)

► **Additional configuration - optional**
Workers per dataset object

Step 23: configure the bounding box tool

After you have configured and verified your instructions, select Create to create the labeling job.

Bounding box labeling tool

Provide labeling instructions with examples below for workers. Workers will be viewing these instructions when they perform your task. Workers can choose up to 50 labels. See guidelines for [See guidelines for creating high-quality instructions](#)

H1 H2 B I A

Good example

Enter description of a correct bounding box label

Add image here

Bad example

Enter description of an incorrect bounding box label

Add image here

Enter a brief description of the task

For each category, select the category from the list and draw a box around each object that this category applies too.

Labels

Add up to 50 labels

×

×

You can add 48 more labels.

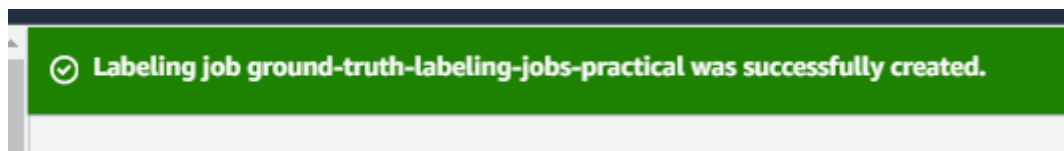
► Additional instructions - optional

Cancel

Previous

Create

Step 24: Monitoring Your Labeling Job



Labeling jobs Info			
<input type="text" value="Search labeling jobs"/>			
	Name	Status	Task type
<input type="radio"/>	ground-truth-labeling-jobs-practical	<input checked="" type="radio"/> In progress	Bounding box