

1. Sign in to the AWS and select Macie from security identity and compliance and choose the AWS Region where you want to start.

2. Choose “**Get started**”.

Get started with Macie

Automatically discover sensitive data across all of your organization's S3 buckets.

Review detailed findings to take remediation action.

Get started

3. Choose “**Enable**” Macie.

Enable Amazon Macie

When you enable Macie, Macie automatically creates a service-linked role for your account. This role gives Macie the permissions that it needs to perform tasks such as gather information about the data that you store in Amazon S3, evaluate and monitor your S3 buckets for security and access control, and run sensitive data discovery jobs that you create to find and report sensitive data in the buckets.

[View role permissions](#)

After you enable Macie, Macie gathers information about your buckets, such as the storage size, encryption settings, and public access settings for each bucket. Macie also begins monitoring the buckets for security and access control, notifying you if the security of a bucket is reduced in some way. You can evaluate this feature at no charge for the first 30 days, and review estimated costs before charges begin to accrue.

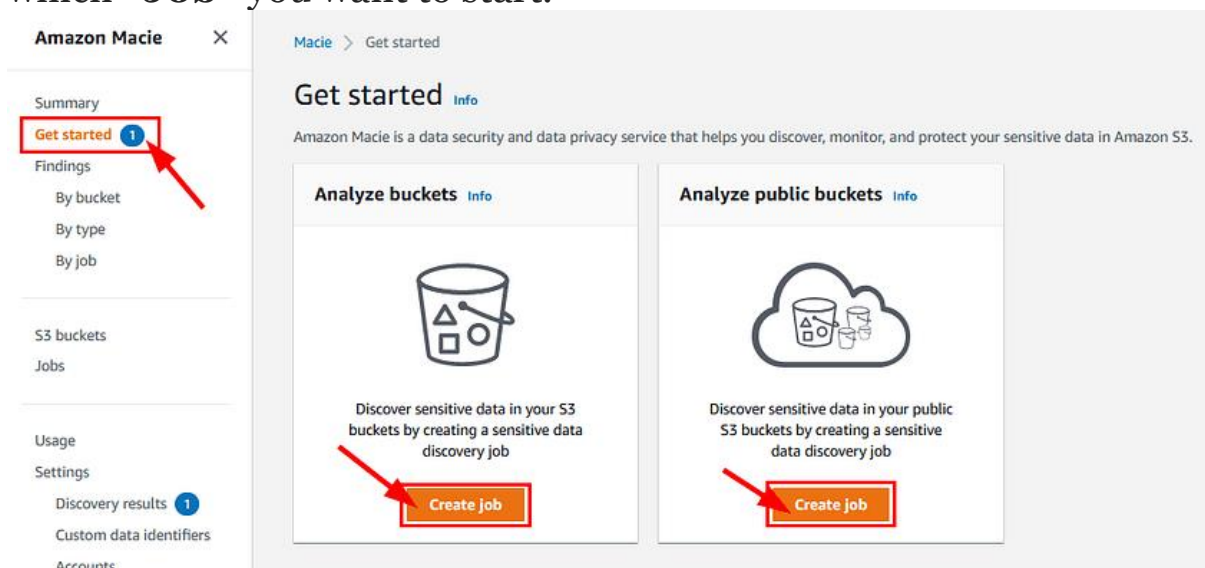
To discover sensitive data, create and configure sensitive data discovery jobs to analyze data in buckets that you specify. There's no charge for analyzing up to 1 GB of data each month. For more information, see [Amazon Macie pricing](#) [\[2\]](#)

Cancel

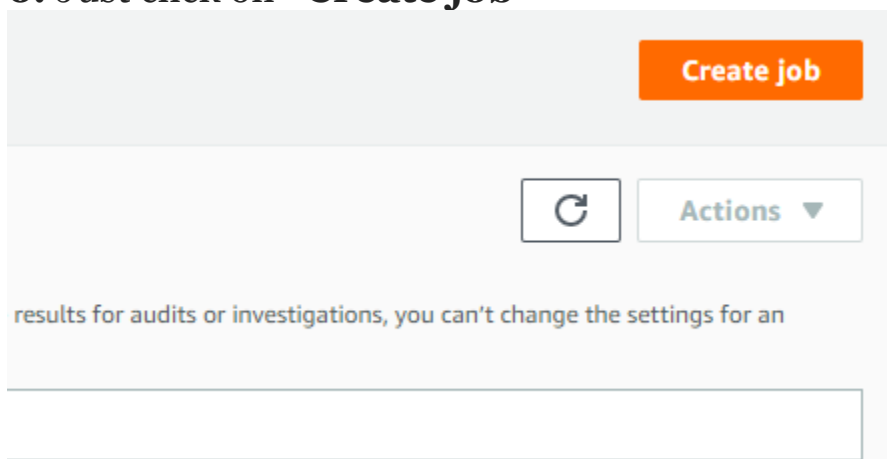
Enable Macie

4. Before enabling the “**Job**” we need to have data for analysis which need to be stored the data in S3. Store some data in S3 by creating a bucket and upload objects and needed permissions according to your consideration.

5. And then just click on “**Get started**” in the menu and select which “**Job**” you want to start.



6. Just click on “**Create job**”



7. Click on the criteria you desire “select specific buckets” or “select specific bucket criteria”. And choose the buckets you want to monitor manually by clicking the check box. Click next.

Choose S3 buckets [Info](#)

A job can analyze objects in one or more S3 buckets. Specify how you want to choose buckets that contain objects for the job to analyze.

☒ **Select specific buckets**
Manually select each bucket that contains objects for the job to analyze. If the job runs more than once, it analyzes objects in the same buckets each time it runs.

☐ **Specify bucket criteria**
Enter criteria that determine which buckets contain objects for the job to analyze. If the job runs more than once, it can analyze objects in different buckets each time it runs, as your bucket inventory changes over time.

Select S3 buckets (2)

This table lists S3 buckets for your account. Select the check box for each bucket to include in the job's analysis.

<input type="checkbox"/>	Bucket	Account	Classifiable obj...	Classifiable s...	Monitored by job	Latest job run
<input type="checkbox"/>	cf-templates-1k71d1mpz01ik-ap-sout...	546010718980	0	0	No	

8. Review the buckets to make sure the needed buckets are added. Click “Next”

Review S3 buckets

Macie and customer managed AWS KMS keys
To analyze objects encrypted with a customer managed AWS KMS key, ensure that Macie is allowed to use the key. [Learn more](#)

S3 buckets (1)
This table lists the S3 buckets that you selected for the job. The estimated cost to analyze a bucket is based on the size and types of objects in the bucket. It assumes that any compressed files use a 3:1 compression ratio.

Bucket name	Account	Classifiable objects	Classifiable si...	Monitored by j...	Estimated c...
	5460107189...	0	0	No	\$ 0.00

9. Choose the scheduled job or one-time job as per your requirement. Here i am selecting a one-time job. In additional settings, We have included or excluded different options. Here I am selecting the storage size and specifying between 50 KB to 200 KB. We have different options here in object criteria and select according to your requirement. Click Next.

Step 2

Review S3 buckets

Step 3

Refine the scope

Step 4

Select managed data identifiers

Step 5

Select custom data identifiers

Step 6

Enter general settings

Step 7

Review and create

Sensitive data discovery options

☒ Scheduled job
Analyze objects on a scheduled frequency

☐ One-time job
Analyze existing objects one time only

Update frequency

Daily

☒ Include existing objects
Select this option to analyze new and existing objects. To analyze only new objects, clear this option.

Sampling depth

100

%

Sample a subset of objects based on depth percentage

▶ Additional settings

Cancel

Previous

Next

▼ Additional settings

Enter criteria that determine which objects to include or exclude from the job's analysis. If you don't enter any criteria, the job analyzes all objects in the buckets.

Object criteria

Storage size

Larger than

KB

Smaller than

KB

Include

Exclude

Include

Storage size : Larger than 50 KB. Smaller than 200 KB.

Delete

Exclude

You haven't entered any exclusion criteria yet.

For reference, I am adding the screenshot of the size of the file i uploaded in S3 bucket.

AWS Region

Asia Pacific (Mumbai) ap-south-1

Last modified

August 24, 2022, 14:50:35 (UTC+05:30)


Size

178.8 KB

Type

pdf

Key

 Maice.pdf

10. Select manage data identifiers. Click next.

Select managed data identifiers [Info](#)

A managed data identifier is a set of built-in criteria that detects a specific type of sensitive data. Select the managed data identifiers for the job to use.

Managed data identifier options

Select the managed data identifiers to use.

Selection type

☒ All

Use all managed data identifiers.

☐ Exclude

Use all managed data identifiers except specific ones that you select.

☐ Include

Use only specific managed data identifiers that you select.

☐ None



Don't use any managed data identifiers.

11. To create custom identifiers for better scanning we need to click on manage custom identifiers.

Select custom data identifiers [Info](#)

A custom data identifier is a set of criteria that you define to detect sensitive data. Select each custom data identifier that you want the job to use.

Custom data identifiers

 [Manage custom identifiers](#) 

<input type="checkbox"/>	Identifier name	Description
You haven't created any custom data identifiers yet.		

Cancel


Previous

Next

12. It will redirect to these page. And click on create. Here we are customizing it in regex expressions by entering `[0-9]` expression any data present in the object. And customizing severity based on the number you prefer. Click on create.

[Macie](#) > Custom data identifiers

Custom data identifiers (0) [Info](#)



Actions ▼

[Create](#)

A custom data identifier is a set of criteria that you define to detect sensitive data.

<input type="checkbox"/>	Name	Description
--------------------------	------	-------------

New custom data identifier [Info](#)

We recommend that you never enter sensitive data in the name or description of an identifier. Other users might be able to see the data in these fields.

Name

Provide name as per user

Description - *optional*

This identifier helps you to scan the bucket or object customized.



Regular expression

Enter the regular expression (regex) that defines the pattern to match.

[0-9]



Keywords - *optional*

Enter up to 50 keywords, separated by commas (,) that define specific text to match. Macie includes a result for text that contains any of these keywords, if the result matches the regex pattern and is within distance of one of these words.

Maximum match distance - optional

Enter the maximum allowable distance between text that matches the regex pattern and the keywords. The default value is 50. Macie includes or excludes a result based on the proximity of a keyword to text that matches the regex pattern.

Default: 50

Severity

Finding severity is based on the number of occurrences of text that matches the preceding criteria.

- ☐ Use Medium severity for any number of matches (default)
- ☒ Use custom settings to determine severity

Occurrences threshold		Severity level		
<input type="text" value="1"/>	or more	<input type="text" value="Low"/>	▼	<input type="button" value="Remove"/>
<input type="text" value="5"/>	or more	<input type="text" value="Medium"/>	▼	<input type="button" value="Remove"/>
<input type="text" value="11"/>	or more	<input type="text" value="High"/>	▼	<input type="button" value="Remove"/>

We can find the Custom data identifiers. We created before.

Custom data identifiers (1) [Info](#)



A custom data identifier is a set of criteria that you define to detect sensitive data.

<input type="checkbox"/>	Name	Description
<input type="checkbox"/>	Dimplecustom	This identifier helps you to scan the bucket or object custo...

13. Select the job we created for custom identifiers. Click next.

Select custom data identifiers [Info](#)

A custom data identifier is a set of criteria that you define to detect sensitive data. Select each custom data identifier that you want the job to use.

Custom data identifiers			Manage custom identifiers 
<input type="checkbox"/>	Identifier name	Description	
<input checked="" type="checkbox"/>	Provide name	This identifier helps you to scan the bucket or object customiz...	

Cancel Previous **Next**

14. Give a job name and tags click next.

Enter general settings [Info](#)

Enter a name for the job. You can also enter a description and assign tags to the job.

Name and description
<p>Job name</p> <div>Provide name</div>
<p>Job description - optional</p> <div></div>
<p>Tags - optional</p>

15. Review and create it.