

AWS Documentation

Author - Preeti Chatterjee

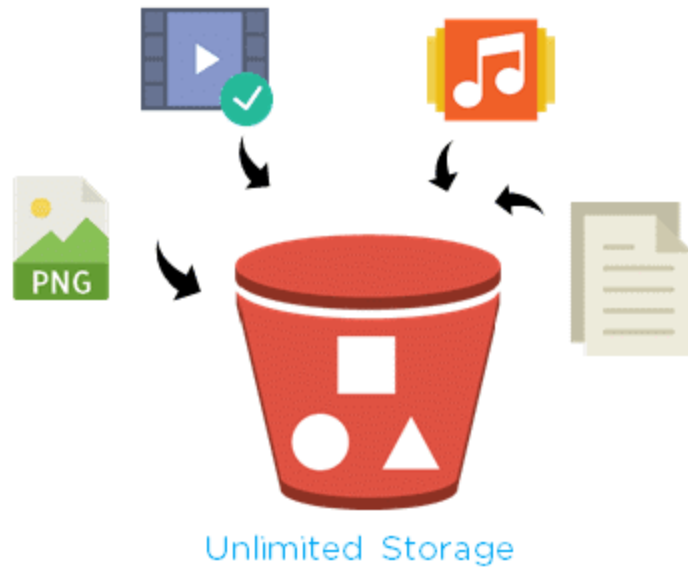
[This is created as a part of my Internship tenure at MinimumQue]

To understand the whole process-overview, we need to familiarise ourselves with a few terms and how they work. The list of such terms are:

- *Amazon S3*
- *Amazon Sagemaker*
- *Amazon Eventbridge*
- *AWS Lambda*

Let's first go through what each of these does:

- **Amazon s3:** Amazon S3 or Amazon Simple Storage Service is a service offered by Amazon Web Services (AWS) that provides object storage through a web service interface. Amazon S3 uses the same scalable storage infrastructure that Amazon.com uses to run its global e-commerce network. Amazon S3 can be employed to store any type of object, which allows for uses like storage for Internet applications, backup and recovery, disaster recovery, data archives, data lakes for analytics, and hybrid cloud storage.



Whatever training, testing data and model we train we save in sagemaker we'll also save in s3 bucket. Anytime we want to reference it, we can retrieve it because reading from s3 is good since it's scalable and we can hit any number of requests from this instance.

- **Amazon Sagemaker:** Amazon SageMaker is a cloud machine-learning platform that was launched in November 2017. SageMaker enables developers to create, train, and deploy machine-learning models in the cloud. SageMaker also enables developers to deploy ML models on embedded systems and edge-devices.

Here we need to know about two terms:

- ☐ *Notebook Instances*
- ☐ *Lifecycle Configurations*

Notebook Instances:

To create a notebook instance in sagemaker we go to “Create notebook instance” in sagemaker.

After choosing notebook instance name, instance type as ml.t2.medium we create an IAM role.

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, bucket-name-2

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](#)

Cancel

Create role

After doing the above we would have successfully created an IAM role.
Note: We can also use an existing role if we already had one.

Amazon SageMaker provides pre-built fully managed notebook instances that run Jupyter notebooks. The notebook instances include example code for common model training and hosting exercises. [Learn more](#)

Notebook instance settings

Notebook instance name

Preeti-internship

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.t2.medium

Elastic Inference [Learn more](#)

none

► Additional configuration

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-20210728T133284



Success! You created an IAM role.

[AmazonSageMaker-ExecutionRole-20210728T133284](#)



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

Then, we create a notebook instance and when the status is “**InService**” and open jupyter following that. **Our Notebook Instance is ready to be used now.**

Lifecycle Configurations:

Let us consider a scenario where we have scheduled our sagemaker notebook instance to run at a scheduled time. Now, since the notebook

instance started we don't want it to sit idle. So, we need some script to run as soon as the sagemaker instance starts.

Scripts

Start notebook**Create notebook**

This script will be run each time an associated notebook instance is started, including during initial creation. If the associated notebook instance is already started, it will be run the next time it is stopped and started. See [a curated list of sample scripts](#)

```
1 set -e
2
3 ENVIRONMENT=python3
4 NOTEBOOK_FILE="/home/ec2-user/SageMaker/script1_mat117.ipynb"
5
6 AUTO_STOP_FILE="/home/ec2-user/SageMaker/auto-stop.py"
7
8 source /home/ec2-user/anaconda3/bin/activate "$ENVIRONMENT"
9
10 jupyter nbconvert "$NOTEBOOK_FILE" --to python --ExecutePreprocessor.kernel_name=python3 --execute
11
12 source /home/ec2-user/anaconda3/bin/deactivate
13
14 # PARAMETERS
15 IDLE_TIME=60 # 1 minute
16
17 echo "Fetching the autostop script"
18 wget https://raw.githubusercontent.com/aws-samples/amazon-sagemaker-notebook-instance-lifecycle-co
```

Cancel**Update**

We need to specify a few terms & concerns about the script here:

- ☐ **NOTEBOOK_FILE** - It runs the notebook file as soon as the sagemaker instance starts
- ☐ **AUTO_STOP_FILE** - This script stops our notebook instance without any manual intervention after **IDLE_TIME**
- ☐ **IDLE_TIME** - Parameter which decides how much time (in seconds) the sagemaker instance sits idle before it stops
- ☐ The more time our sagemaker instance runs, the more cost we incur

Studio Notebooks			
On-Demand Notebook Instances			
Processing			
Data Wrangler			
Feature Store			
Training			
Real-Time Inference			
Batch Transform			
Amazon SageMaker Studio Notebooks Amazon SageMaker Studio Notebooks are one-click Jupyter notebooks that can be spun up quickly. The underlying compute resources are fully elastic and the notebooks can be easily shared with others enabling seamless collaboration. You are charged for the instance type you choose, based on the duration of use.			
Region: US West (Oregon) ▾			
Standard Instances	vCPU	Memory	Price per Hour
ml.t3.medium	2	4 GiB	\$0.05
ml.t3.large	2	8 GiB	\$0.10
ml.t3.xlarge	4	16 GiB	\$0.20
ml.t3.2xlarge	8	32 GiB	\$0.399
ml.m5.large	2	8 GiB	\$0.115
ml.m5.xlarge	4	16 GiB	\$0.23
ml.m5.2xlarge	8	32 GiB	\$0.461
ml.m5.4xlarge	16	64 GiB	\$0.922
ml.m5.8xlarge	32	128 GiB	\$1.843
ml.m5.12xlarge	48	192 GiB	\$2.765
ml.m5.16xlarge	64	256 GiB	\$3.686
ml.m5.24xlarge	96	384 GiB	\$5.53

- **Amazon Eventbridge:** This helps us schedule any recurring tasks. To operate this, we have to create “Rules” under “Events” and then we *create a*

rule. While creating rules we have two options - either we can choose an **event pattern** or set a **schedule** to invoke targets. Event pattern implies that when something occurs then only the task gets triggered, for example we consider data deposits in s3 buckets as the trigger event. On the other hand, if we need our task to run at certain intervals of time using “Fixed rate every” or at a fixed time everyday using “Cron expression” [Image given below for reference], that would imply scheduling. **So, in conclusion, Amazon Eventbridge helps us schedule a particular task.**

Define pattern

Build or customize an Event Pattern or set a Schedule to invoke Targets.

☐ Event pattern [Info](#)

Build a pattern to match events

☒ Schedule [Info](#)

Invoke your targets on a schedule

☐ Fixed rate every

Hours

☒ Cron expression

CRON expression have six required fields, which are separated by white space. [Learn more about CRON expression.](#) [Enter CRON expression below to see the next 10 trigger date\(s\).](#)

0 15 * * ? *

Next 10 trigger date(s)

Sat, 22 May 2021 15:00:00 GMT
Sun, 23 May 2021 15:00:00 GMT
Mon, 24 May 2021 15:00:00 GMT
Tue, 25 May 2021 15:00:00 GMT
Wed, 26 May 2021 15:00:00 GMT
Thu, 27 May 2021 15:00:00 GMT
Fri, 28 May 2021 15:00:00 GMT
Sat, 29 May 2021 15:00:00 GMT
Sun, 30 May 2021 15:00:00 GMT
Mon, 31 May 2021 15:00:00 GMT

► Sample event(s)

GMT

GMT

Local time zone

- **AWS Lambda:** AWS Lambda is an event-driven, serverless computing platform provided by Amazon as a part of Amazon Web Services. It is a computing service that runs code in response to events and automatically manages the computing resources required by that code.

To create a function, we go to Dashboard under AWS Lambda and click on “Create function”. Then we choose the following and decide the IAM role from the admin.

Lambda > Functions > Create function

Create function [Info](#)

Choose one of the following options to create your function.

Author from scratch



Start with a simple Hello World example.

Use a blueprint



Build a Lambda application from sample code and configuration presets for common use cases.

Container image

Select a container image

Basic information

Function name

Enter a name that describes the purpose of your function.

preeti-internship-sample

Use only letters, numbers, hyphens, or underscores with no spaces.

Runtime [Info](#)

Choose the language to use to write your function. Note that the console code editor supports only Node.js, Python, and Ruby.

Python 3.7

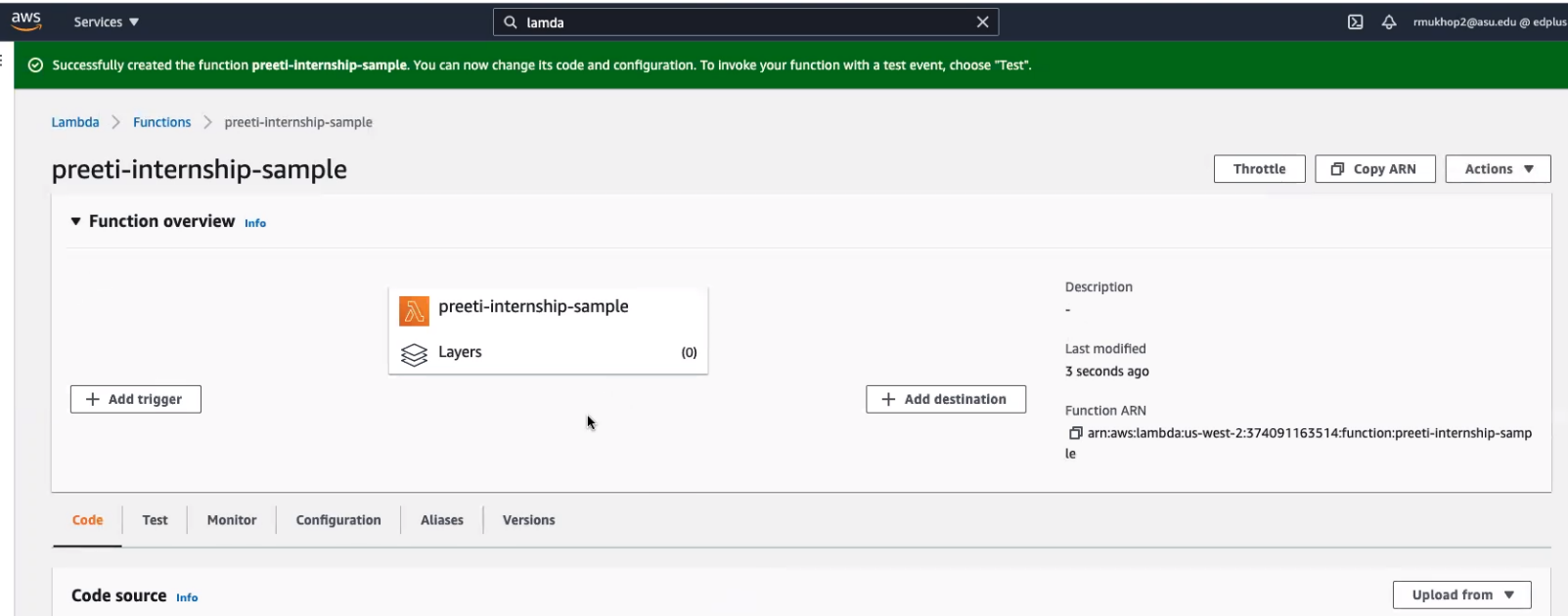
Permissions [Info](#)

By default, Lambda will create an execution role with permissions to upload logs to Amazon CloudWatch Logs. You can customize this default role later when adding triggers.

► Change default execution role

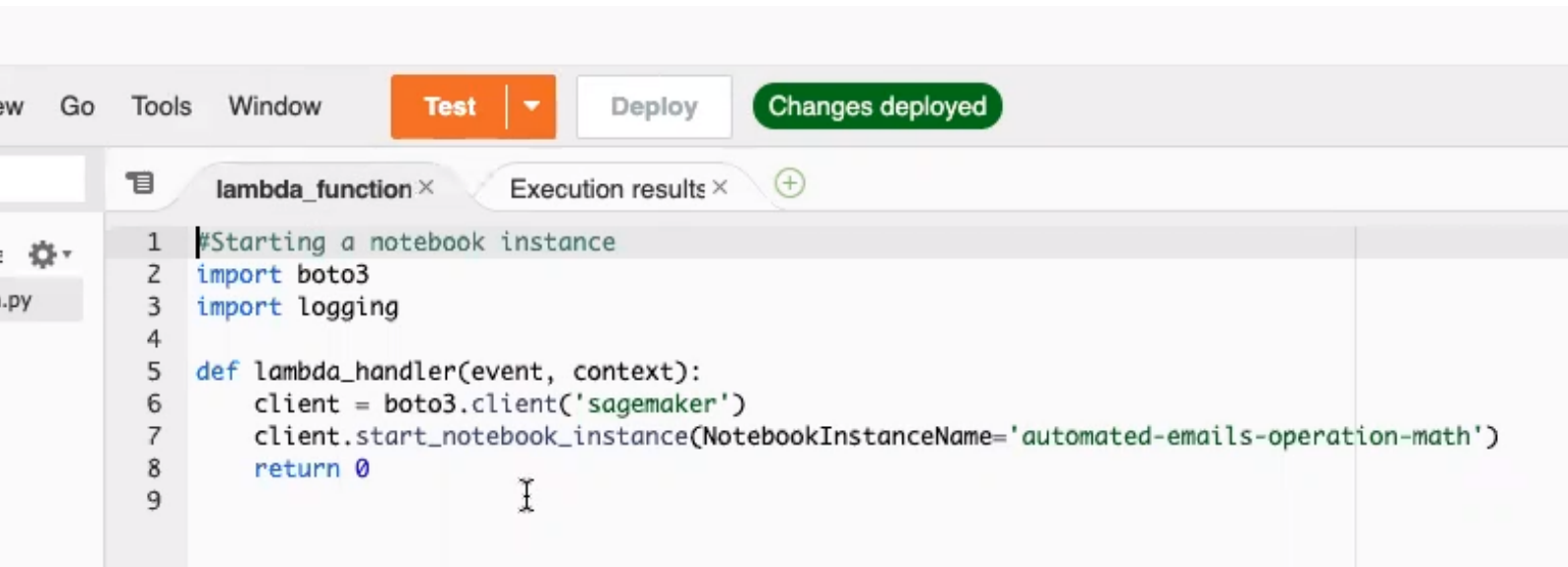
► Advanced settings

This creates our desired lambda function:



Lambda function named “preeti-internship-sample” created

Now, given below is an example code of how to start a notebook instance inside a lambda function:



A lambda function which starts a sagemaker notebook instance named “automated-emails-operation-math”

Bringing Everything Together

- Create a notebook instance
- Configure Lifecycle configurations and link it to notebook instance
- Creating relevant task using AWS Lambda to start the relevant notebook
- Setting a trigger by clicking on “Add trigger” option and choosing Amazon Eventbridge
- Create a new rule or use existing rules to schedule a trigger at a specific time or intervals of time