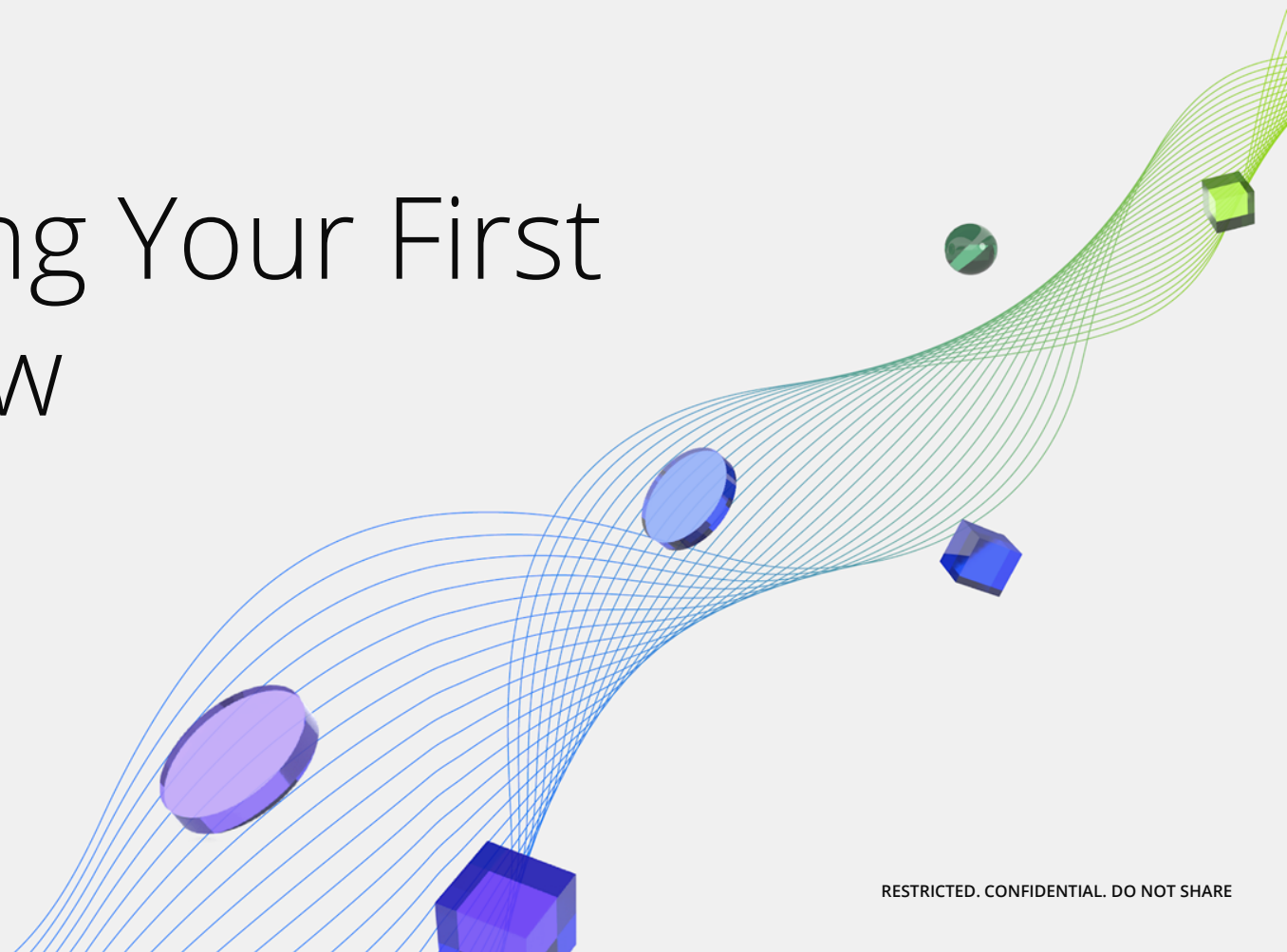
UDACITY

Designing Your First Workflow

Lamia Zain
lamiahasan4@gmail.com
LinkedIn

August, 02, 2024



Connect Sessions | Purpose

A Connect Session **IS**:

- Focused on learning, encouragement & graduation for a group of students coached by a Udacity Session Lead
- Setting weekly study goals
- Helping each other with progress (including peer to peer)
- Keeping everyone accountable for their responsibilities
- A way to meet individuals in tech field & learn about the industry
- **Mandatory**

A Connect Session **IS NOT**:

- A social meetup
- A study group
- A substitute for online learning
- **Optional**

UDACITY

Let’s check your progress

You are encouraged to spend at lest 10 hours/week to graduate.

Presentation date

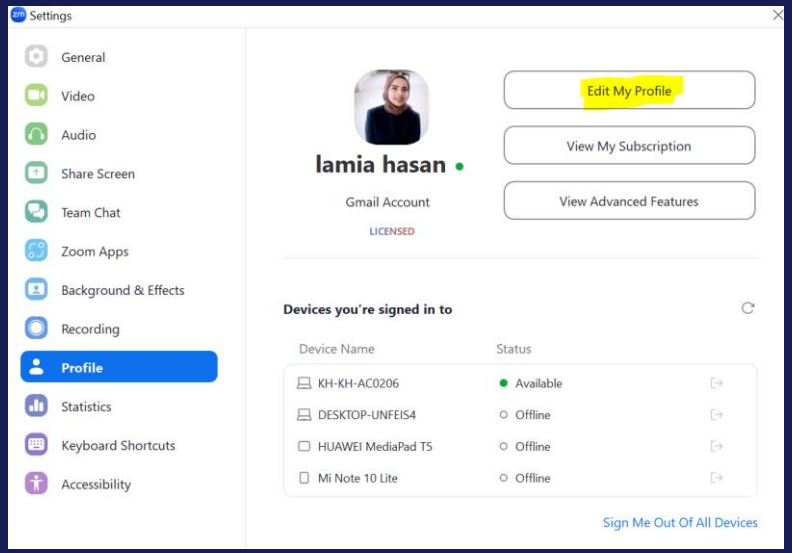
UDACITY

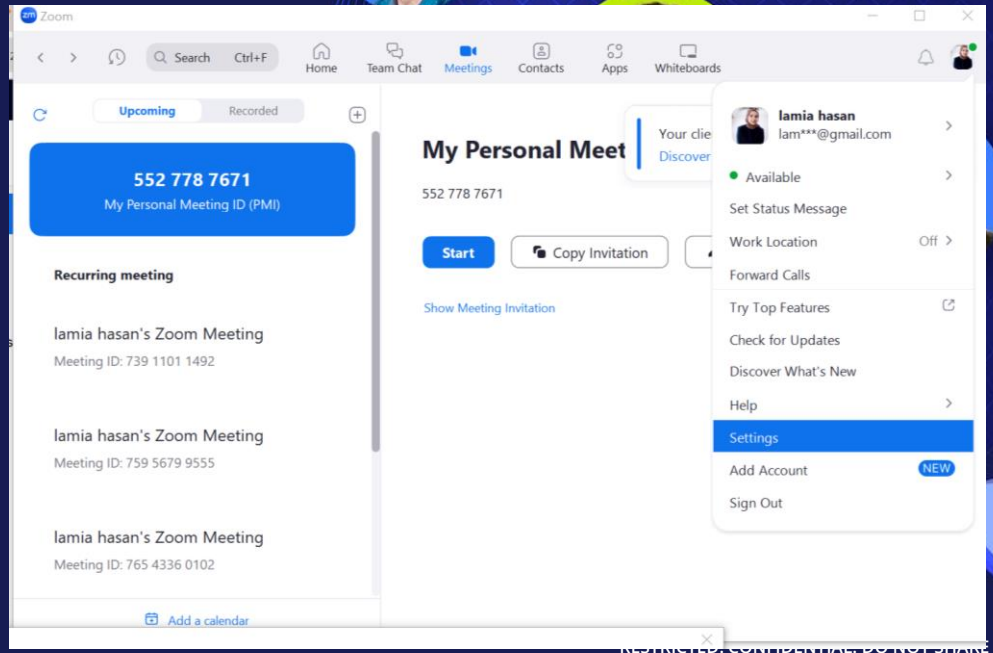
Attendance is taken automatically

Please change your name to be First Name and Last name on Zoom
Like : Lamia Zain

UDACITY

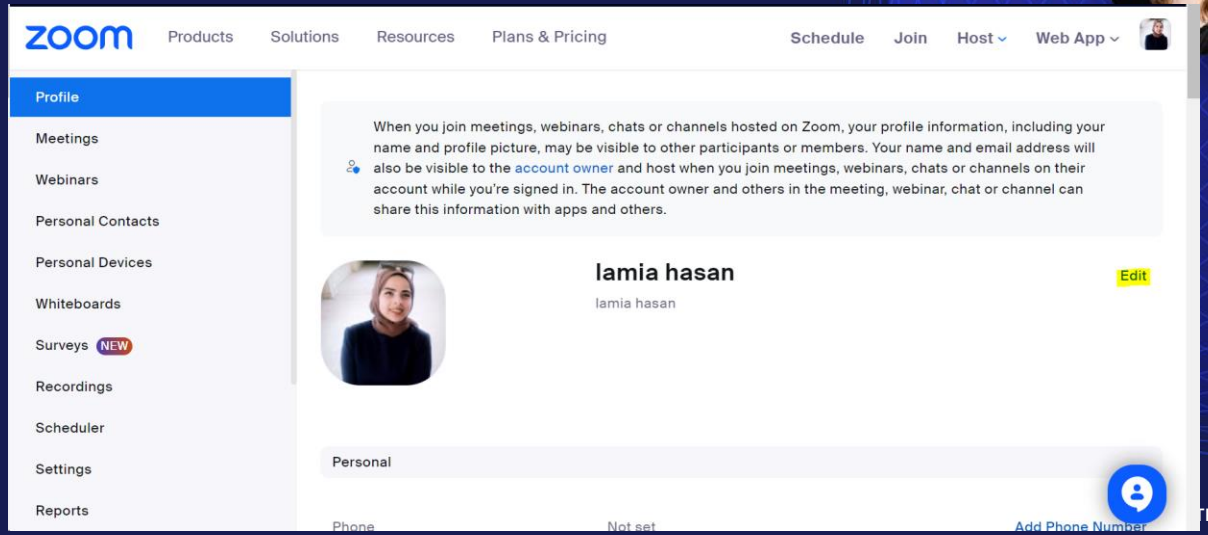
Change your Name on Zoom





UDACITY

Change your Name on Zoom



Session Lead role:

Communication Chart

Issue	Where to go?
Classroom access/ Withdrawal/ Graduation issues/ Plagiarism/ Project Review Inquiries	Email support@udacity.com
Technical Issues, Attendance, Content Related Issues/ Project inquiries	Session Lead
Session Switch/ Community related issues	Community Moderators

Program Timeline & Learning Journey

2024

April | May | June | July | August

- **Program Kickoff** — April 10, 2024
- **Revoking Deadline** — May 12, 2024
- **First Project Submission** — May 1, 2024
- **Second Project Submission** — June 5, 2024
- **Third Project Submission** — July 10, 2024
- **Fourth Project Submission** — August 21, 2024
- **End of Program** — August 28, 2024

Program Period — April 10 - August 28, 2024

Four-weeks Agenda, Weekly schedule

Week 15	Jul 17, 2024			Finish the lessons below from **the Developing your First ML Workflow** Introduction to Developing ML Workflows **[Work on/submit the #4 project: Build a ML Workflow For Scones Unlimited On Amazon SageMaker]**	Developing your First ML Workflow Introduction to Developing ML Workflows
Week 16	Jul 24, 2024			Finish the lessons below from **the Developing your First ML Workflow** SageMaker Essentials **[Work on/submit the #4 project: Build a ML Workflow For Scones Unlimited On Amazon SageMaker]**	Developing your First ML Workflow SageMaker Essentials
Week 17	Jul 31, 2024			Finish the lessons below from **the Developing your First ML Workflow** Designing Your First Workflow **[Work on/submit the #4 project: Build a ML Workflow For Scones Unlimited On Amazon SageMaker]**	Developing your First ML Workflow Designing Your First Workflow
Week 18	Aug 7, 2024			Finish the lessons below from **the Developing your First ML Workflow** Monitoring a ML Workflow **[Work on/submit the #4 project: Build a ML Workflow For Scones Unlimited On Amazon SageMaker]**	Developing your First ML Workflow Monitoring a ML Workflow Project Walkthrough: Build a ML Workflow For Scones Unlimited On Amazon SageMaker

Four-weeks Agenda, Weekly schedule

Week 19	Aug 14, 2024	Aug 14, 2024	**Build a ML Workflow For Scones Unlimited On Amazon SageMaker**	Finish the lessons below from **the Developing your First ML Workflow** **[Work on/submit the #4 project: Build a ML Workflow For Scones Unlimited On Amazon SageMaker]**	Project Walkthrough: Build a ML Workflow For Scones Unlimited On Amazon SageMaker
Week 20	Aug 21, 2024			Prepare any questions you have about the content	Ask me Anything Session
Week 21	**Aug 28, 2024**	**(FINISH & GRADUATE)**			

Student Milestone | Revoking

REVOKING

Revoking is the process by which Udacity removes a student from a Nanodegree program.

AWS reserves the right to revoke you from the program if you do not comply with program requirements.

CRITERIA

Students can be revoked if they fail to:

- Submit Project 1
- Complete the required concepts

The date of **Revoking** for the scholarship is **May 12, 2024**

Code of Conduct | Plagiarism

BASIC RULES

- Project submissions must consist of original work
- Submitted projects will be scanned for plagiarism
- Students who are found to have plagiarised will risk their Nanodegree being revoked
- Read the honor code and the rubric carefully for all projects

Objectives:

1- Processing Jobs using Python SDK.
2- Create Lambda Functions SageMaker Python SDK.
3- Create Lambda Functions SageMaker Console.
4- Invoke lambda functions through a test event
5- Invoke lambda functions from a jupyter notebook.
6- Invoke lambda functions with CloudWatch event rules (Readings).

Processing Job

Processing Jobs are SageMaker service for data preprocessing.
Session Exercise: Create a SklearnProcessor object to drop one column of the Breast Cancer dataset Before training the model.
Documentation:
1- SageMaker SDK
2- Example here

Lambda Functions

Why?

2- No Resources Selection needed (Instance type and count)
1- Small tasks done repeatedly.

You Can:
1- Specify the execution time for lamda function
2- Specify the size of the lambda function
3- Specify Lambda Function policies attached to its IAM role.

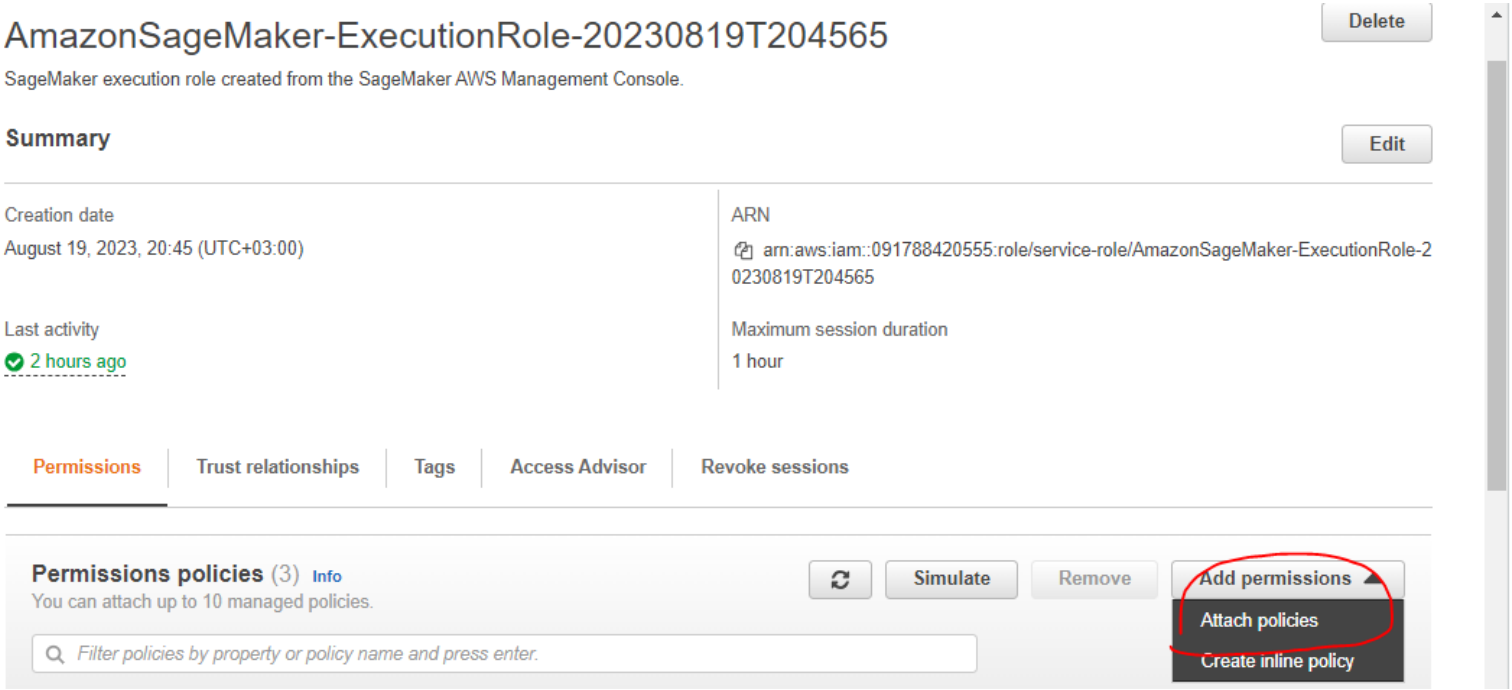
Creating a Lambda Function BOTO3 SDK

Add Lambda Full access permission to the current notebook

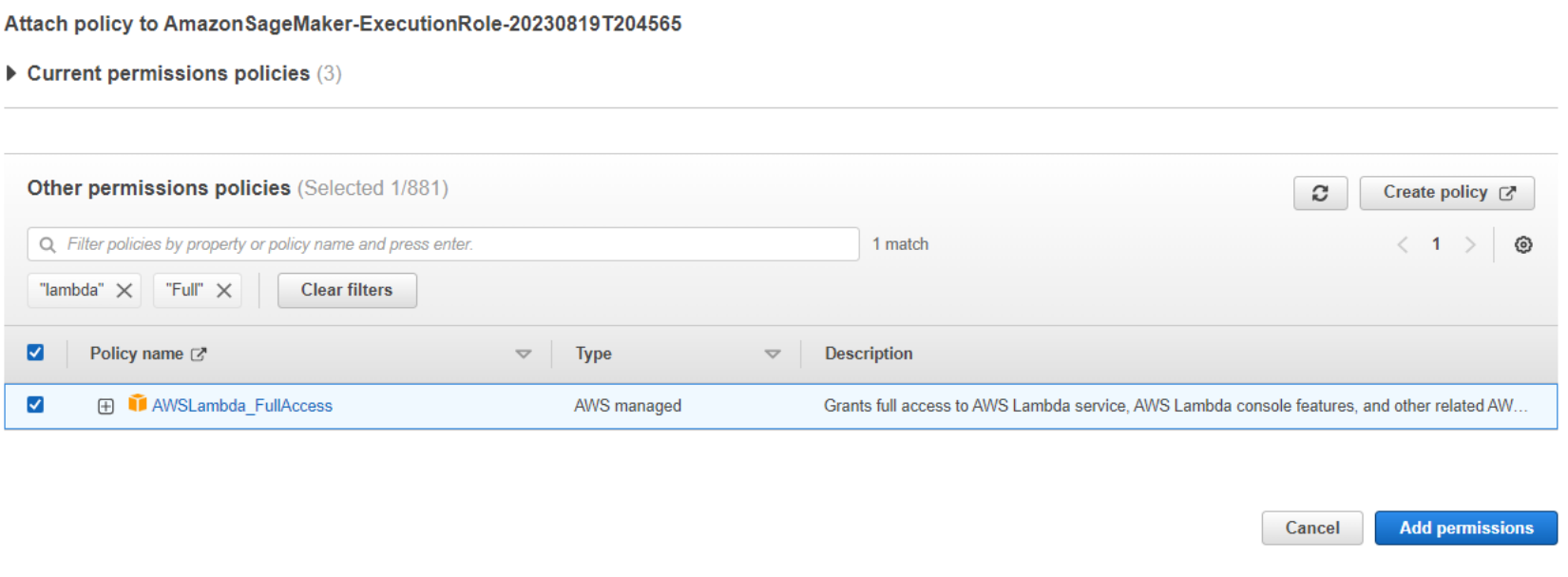
In [2]:
```
import boto3
from sagemaker import get_execution_role

#Make sure you have full access to lambda function

role = get_execution_role()
print(role)

client = boto3.client('lambda') # is used to create a client object for interacting with AWS Lambda,
```

```
arn:aws:iam::091788420555:role/service-role/AmazonSageMaker-ExecutionRole-20230819T204565
```

Go to role in IAM role



Search for **AWSLambdaFullAceess** Permission and add it



boto3.client('lambda').create_function()

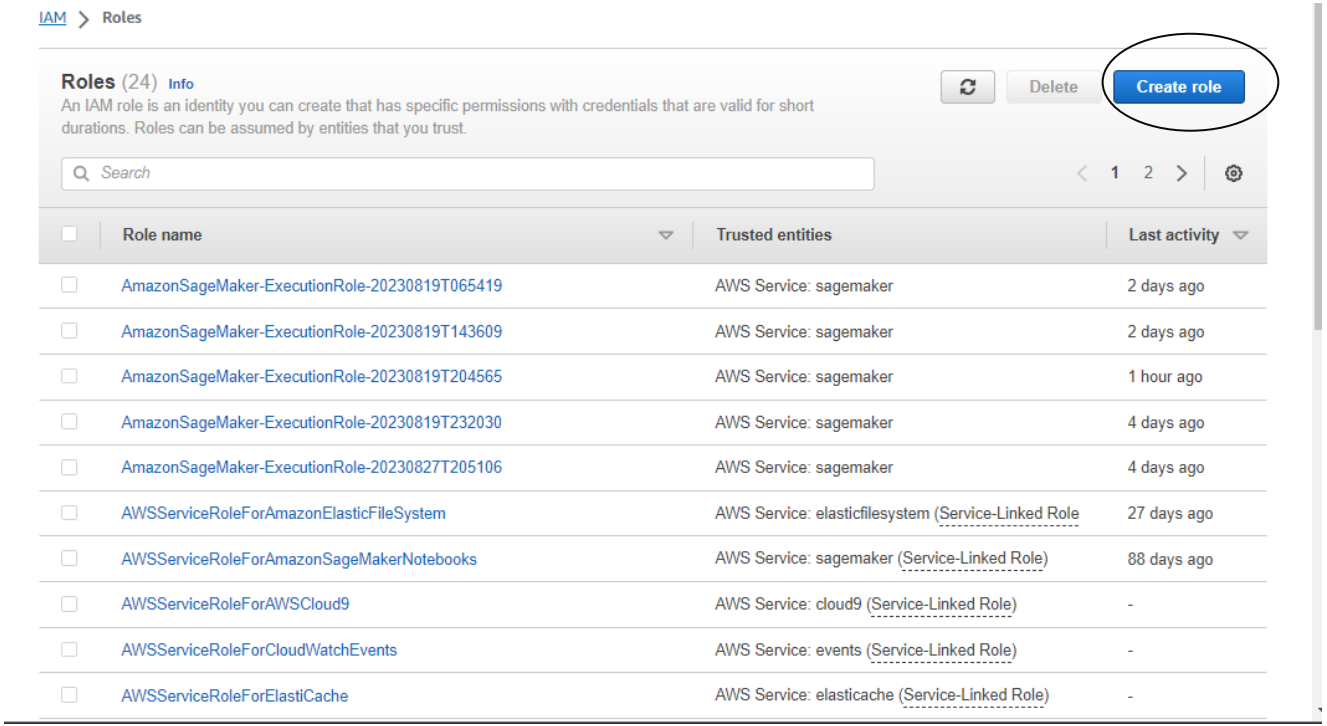
Request Syntax

```
response = client.create_function(
    FunctionName='string',
    Runtime='nodejs'|'nodejs4.3'|'nodejs6.10'|'nodejs8.10'|'nodejs10.x'|'nodejs12.x'|'nodejs14.x'
    Role='string',
    Handler='string',
    Code={
        'ZipFile': b'bytes',
        'S3Bucket': 'string',
        'S3Key': 'string',
        'S3ObjectVersion': 'string',
        'ImageUri': 'string'
    },
    Description='string',
    Timeout=123,
    MemorySize=123,
    Publish=True|False,
    VpcConfig={
        'SubnetIds': [
            'string',
        ],
        'SecurityGroupIds': [
            'string',
        ]
    },
    PackageType='Zip'|'Image',
    DeadLetterConfig={
        'TargetArn': 'string'
    },
    Environment={
```

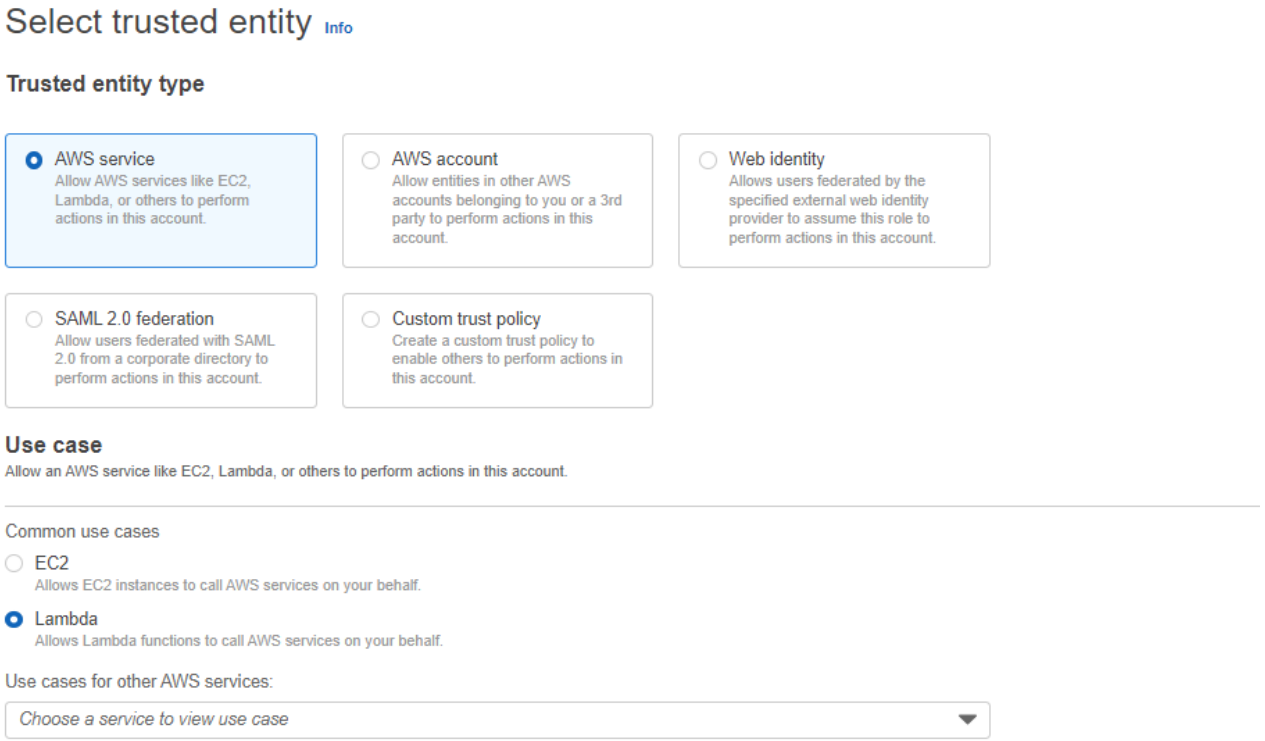
Create a role that access lambda function.

In [4]:
```
response = client.create_function(
    FunctionName ='Lambda_Function2',  #Unique FName
    #The identifier of the function's runtime.Runtime is required if the deployment package is a .zip file archive.
    Runtime = 'python3.10',


    Handler = 'lambda_function.lambda_handler',
    Code = {'ZipFile': b_code},
    Timeout=30,
    MemorySize=1024,
    Publish=True,#Set to true to publish the first version of the function during creation
    #The type of deployment package. Set to Image for container image and set to Zip for .zip file archive.
    PackageType='Zip',
    #The Amazon Resource Name (ARN) of the function's execution role.
    Role = 'arn:aws:iam::091788420555:role/Lambda-Basic-Role'
)
```

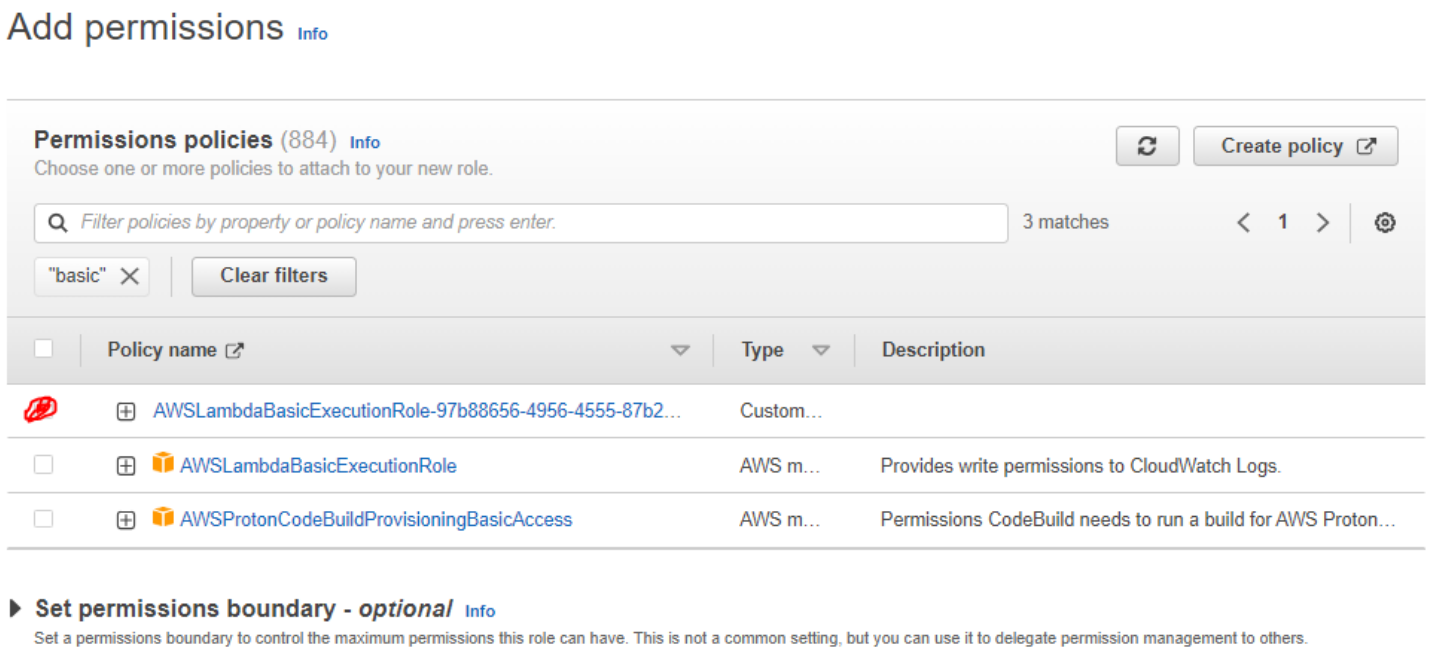
Create a role that access lambda function.



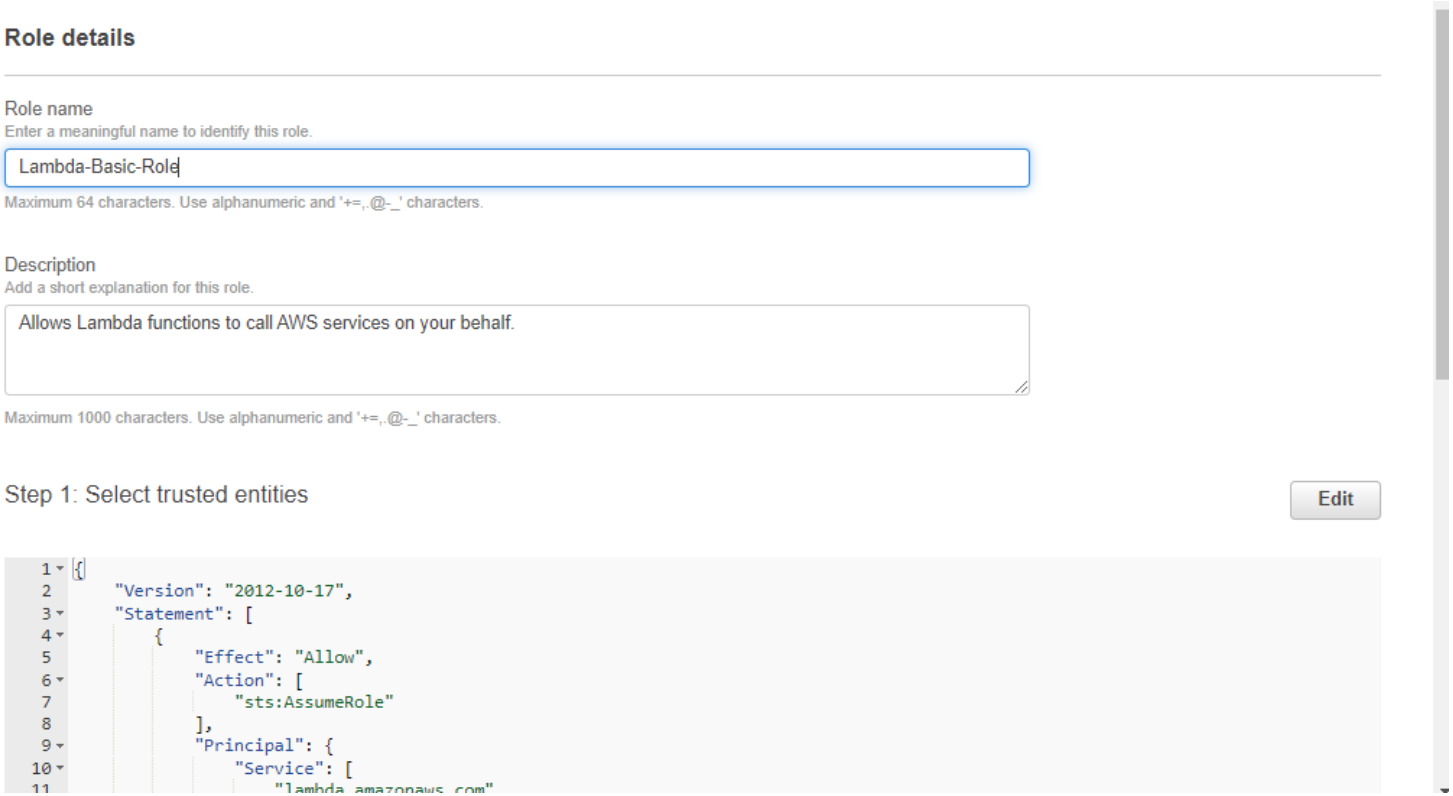
Create a role that access lambda function.



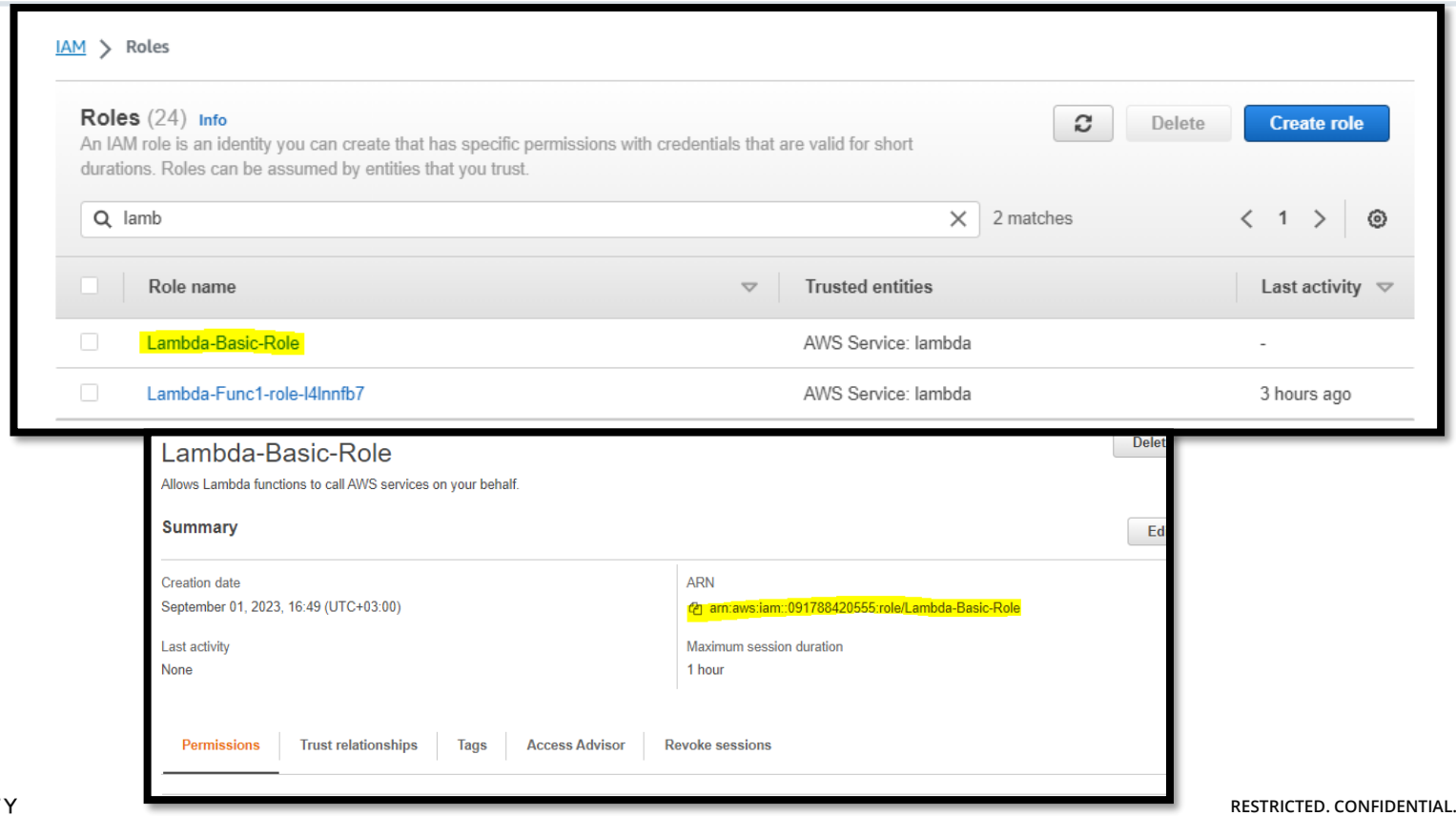
Add lambda basic execution role to policies



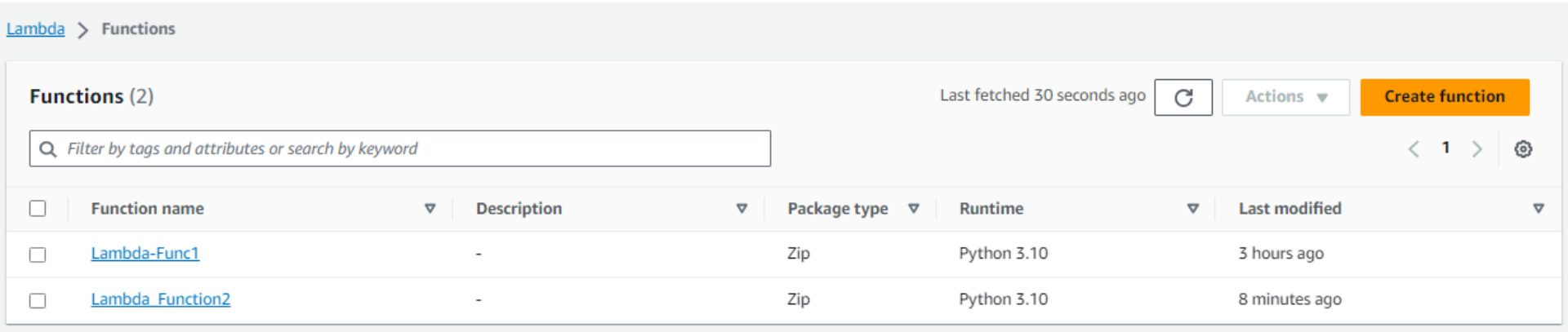
Provide Role Name



Search for the role, copy its ARN



Lambda function is created



Creating a Lambda Function SageMaker Console

Break (10 minutes)

Satisfaction Survey

Exercise#1 Invoking Lambda function Preprocessing Toys' review dartaset

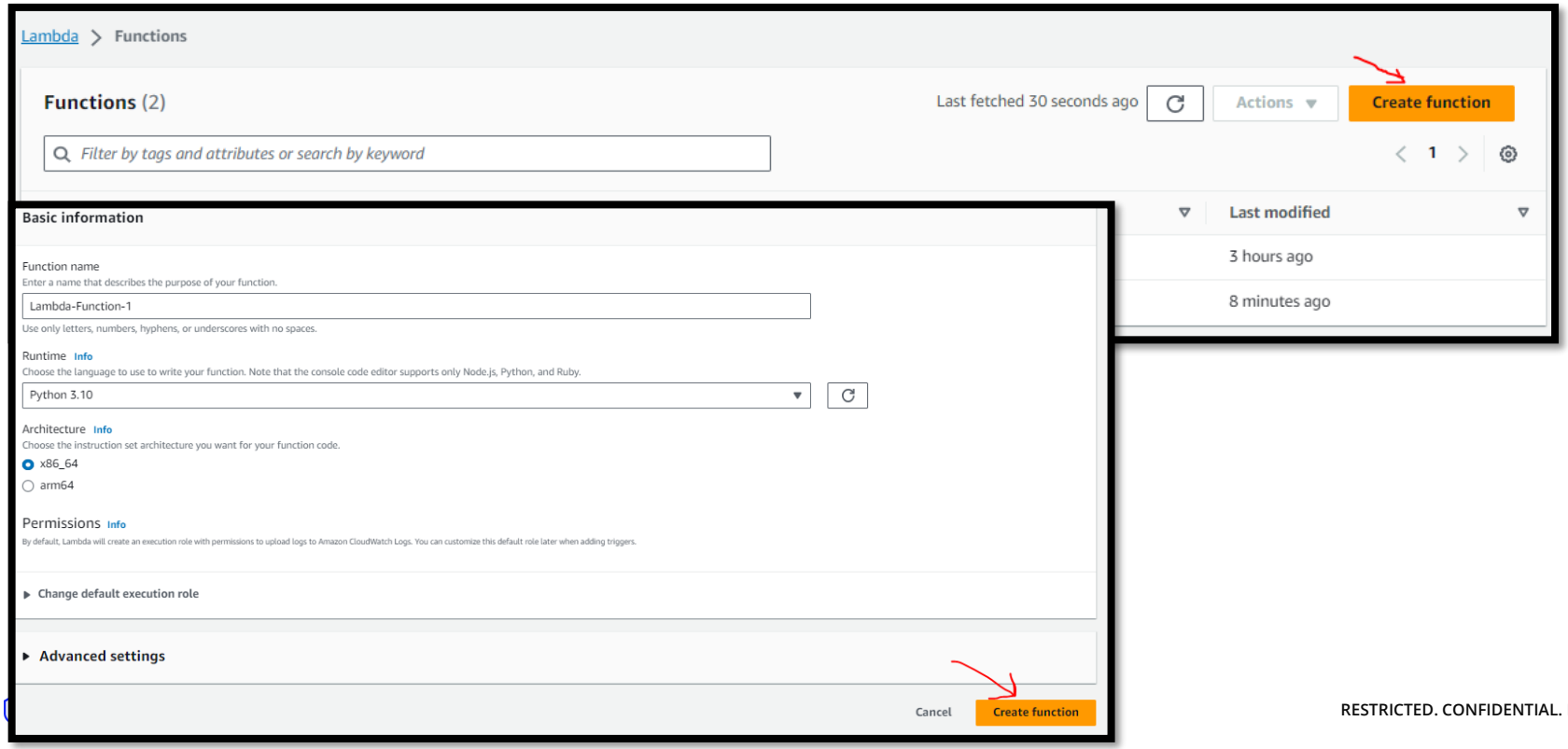
In this exercise, Our goal is:
Use Lambda function to extract Toys' reviews data from a .zip file stored in S3 Bucket, preprocess data to be in this format, split files to training and testing files, and upload to S3 again.

__label__1 linux ready for prime time , intel says , despite all the linux hype , the open-source movement has yet to make a huge splash in the desktop market . that may be about to change , thanks to chipmaking giant intel corp .

__label__2 bowled by the slower one again , kolkata , november 14 the past caught up with sourav ganguly as the indian skippers return to international cricket was short lived .

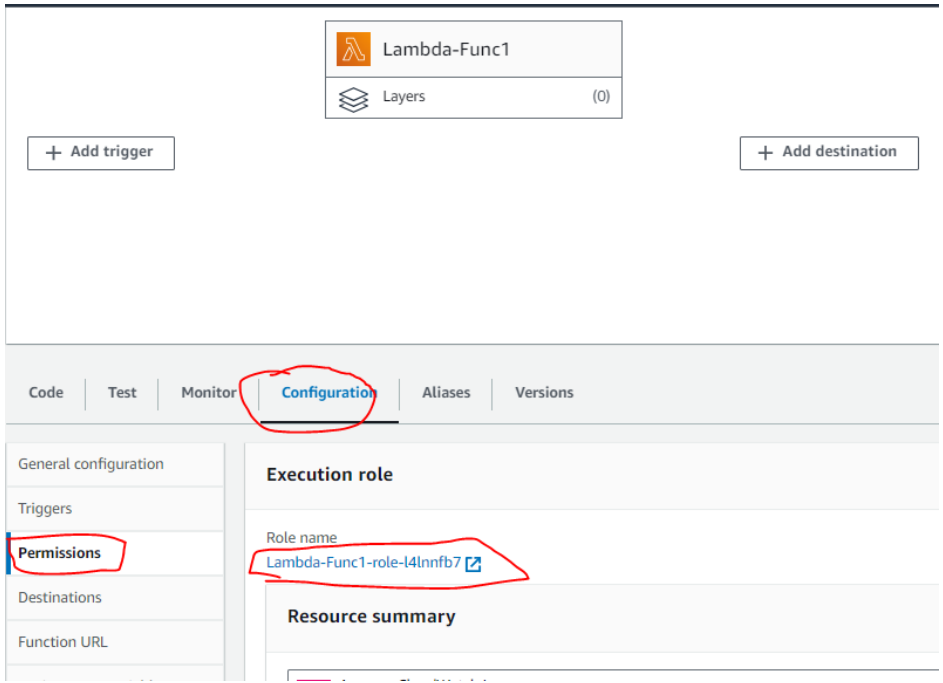
Steps:

- Create Lambda function:



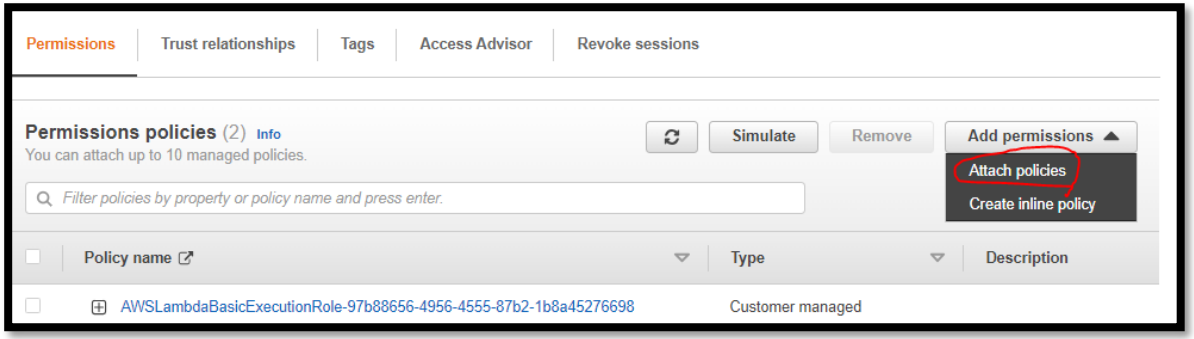
Steps:

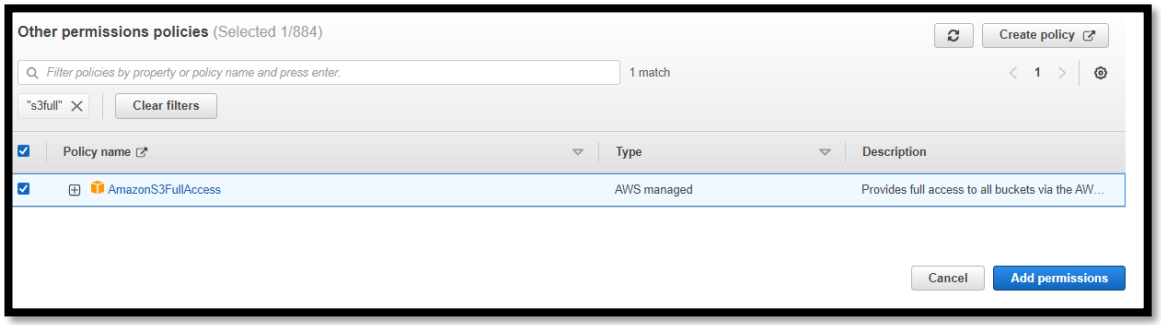
- Go to the corresponding role



Steps:

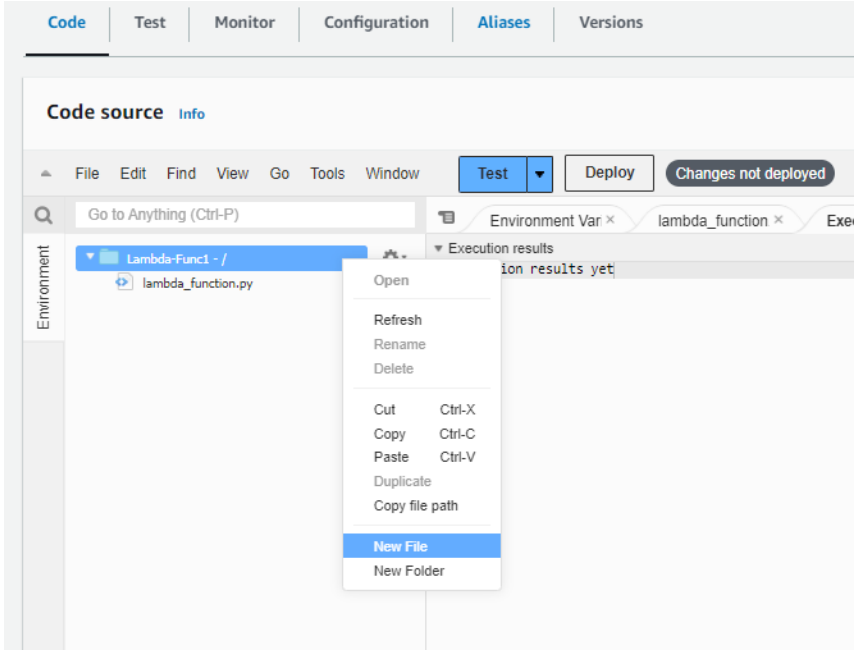
- Add permission to the lambda function to Fully access s3 Bucket.

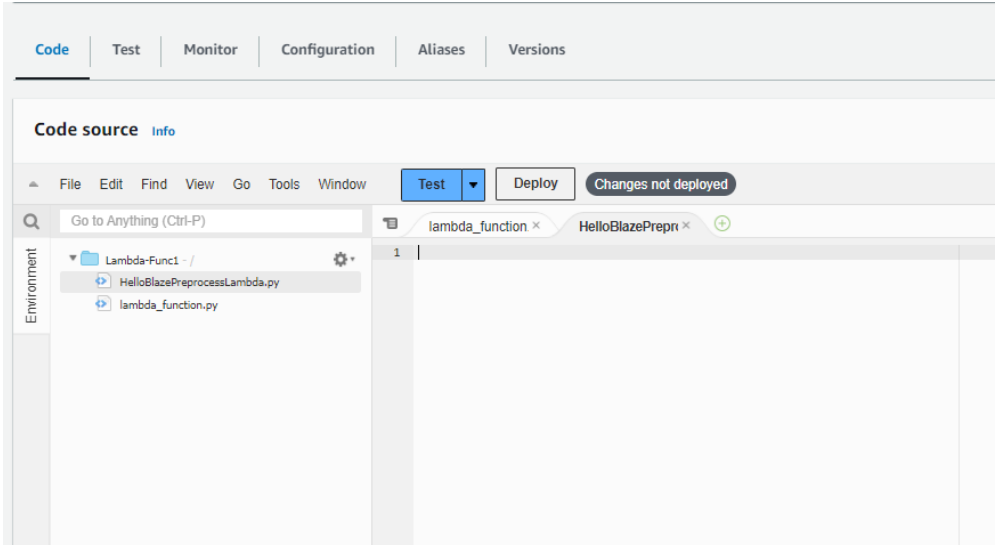




Steps:

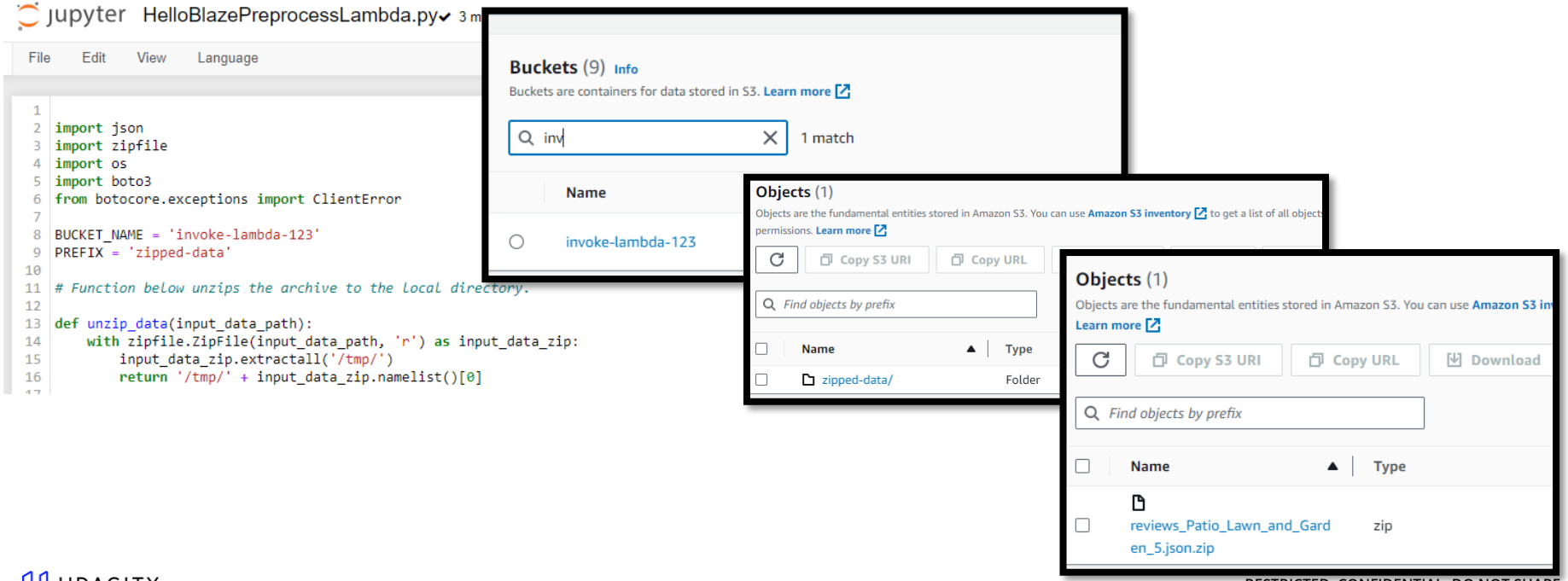
- Create a new python file and name it **HelloBlazePreprocessLambda.py**





Steps:

- Edit the provided **HelloBlazePreprocessLambda.py** to contain valid Bucket and prefix



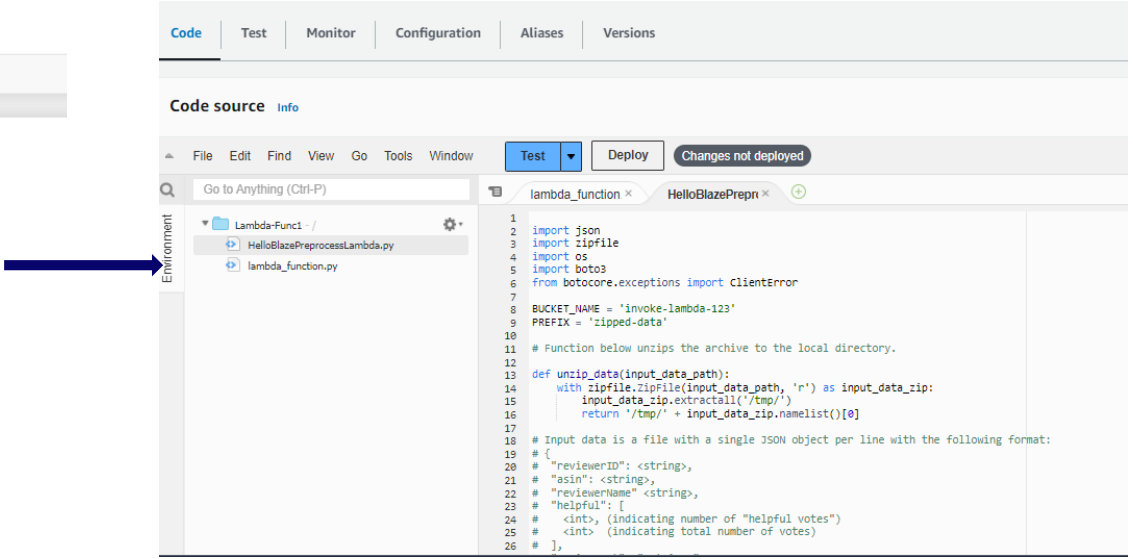
Steps:

- Copy file content to lambda **HelloBlazePreprocessLambda.py** file

jupyter HelloBlazePreprocessLambda.py ✔ 3 minutes ago

File Edit View Language

```
1
2  import json
3  import zipfile
4  import os
5  import boto3
6  from botocore.exceptions import ClientError
7
8  BUCKET_NAME = 'invoke-lambda-123'
9  PREFIX = 'zipped-data'
10
11 # Function below unzips the archive to the local directory.
12
13 def unzip_data(input_data_path):
14     with zipfile.ZipFile(input_data_path, 'r') as input_data_zip:
15         input_data_zip.extractall('/tmp/')
16         return '/tmp/' + input_data_zip.namelist()[0]
```



Steps:

- You can also Zip the two files.py and upload them to lambda at once

Zipping HelloBlazePreprocessLambda.py file

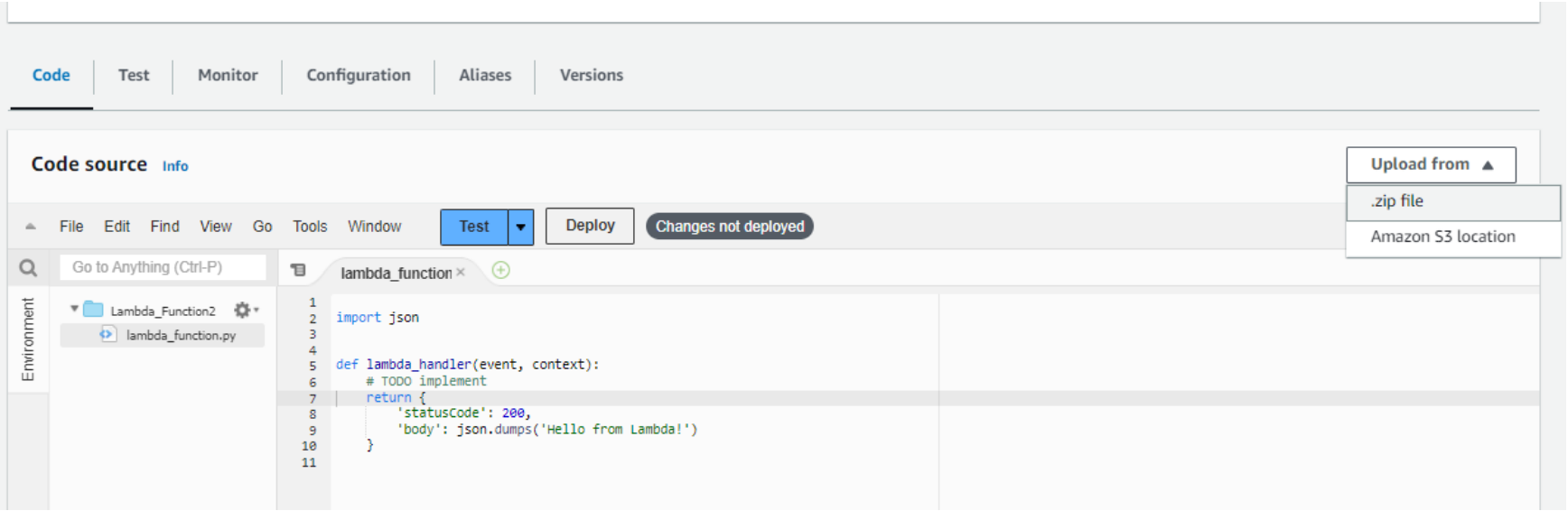
In [1]:
```
%%writefile lambda_function.py

import json

from HelloBlazePreprocessLambda import preprocess

def lambda_handler(event, context):
    # TODO implement
    preprocess(event["s3-dataset-uri"])
    return {
        'statusCode': 200,
        'body': json.dumps('Hello from Lambda!')
    }
```

```
Overwriting lambda_function.py
```

In [2]:
```
from zipfile import ZipFile
import zipfile
# Create a ZIP archive and add the specified files to it

# Specify the name of the file you want to zip
file_to_zip = 'HelloBlazePreprocessLambda.py'

# Specify the name of the ZIP file you want to create

with ZipFile('code.zip','w') as f:
    f.write('lambda_function.py')
    f.write('HelloBlazePreprocessLambda.py')
```

Steps:

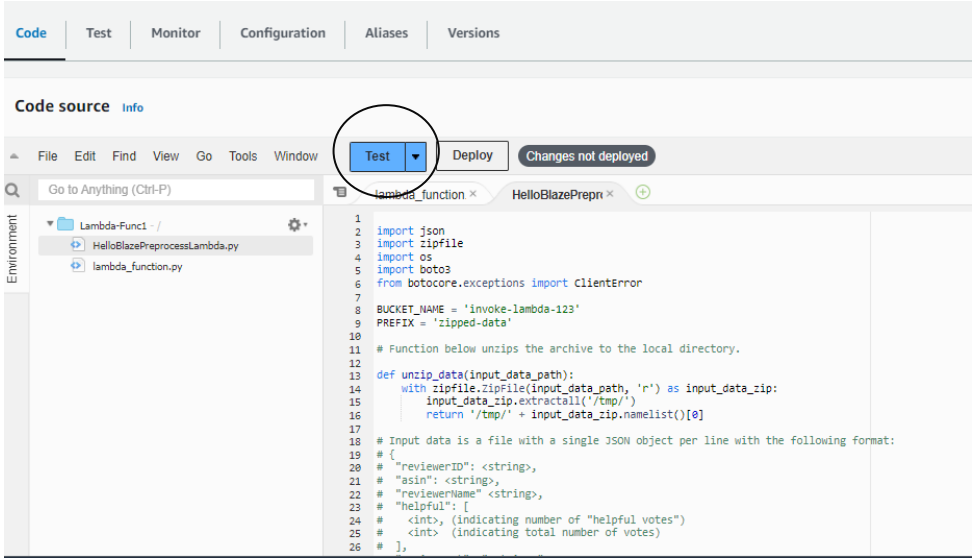
- You can also Zip the two files.py and upload them to lambda at once

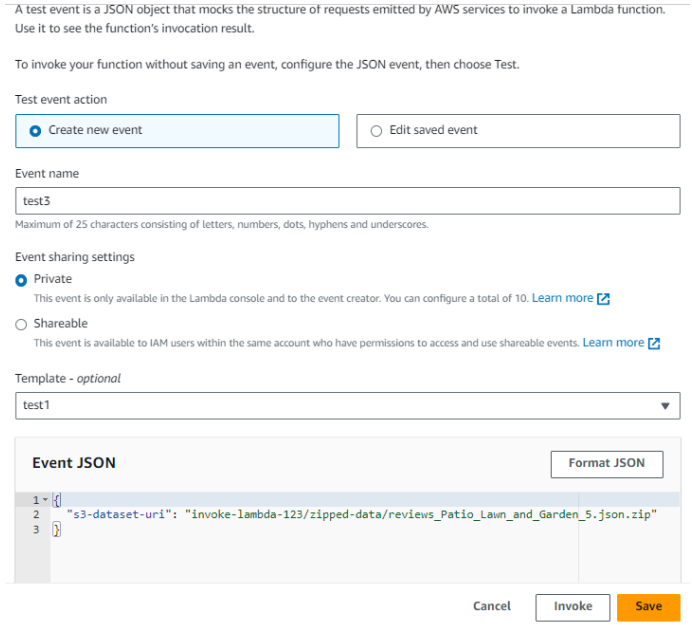


Steps:

- Configure a test event.
- Json Object will be like:

{ "s3-dataset-uri": "invoke-lambda-123/zipped-data/reviews_Patio_Lawn_and_Garden_5.json.zip"}





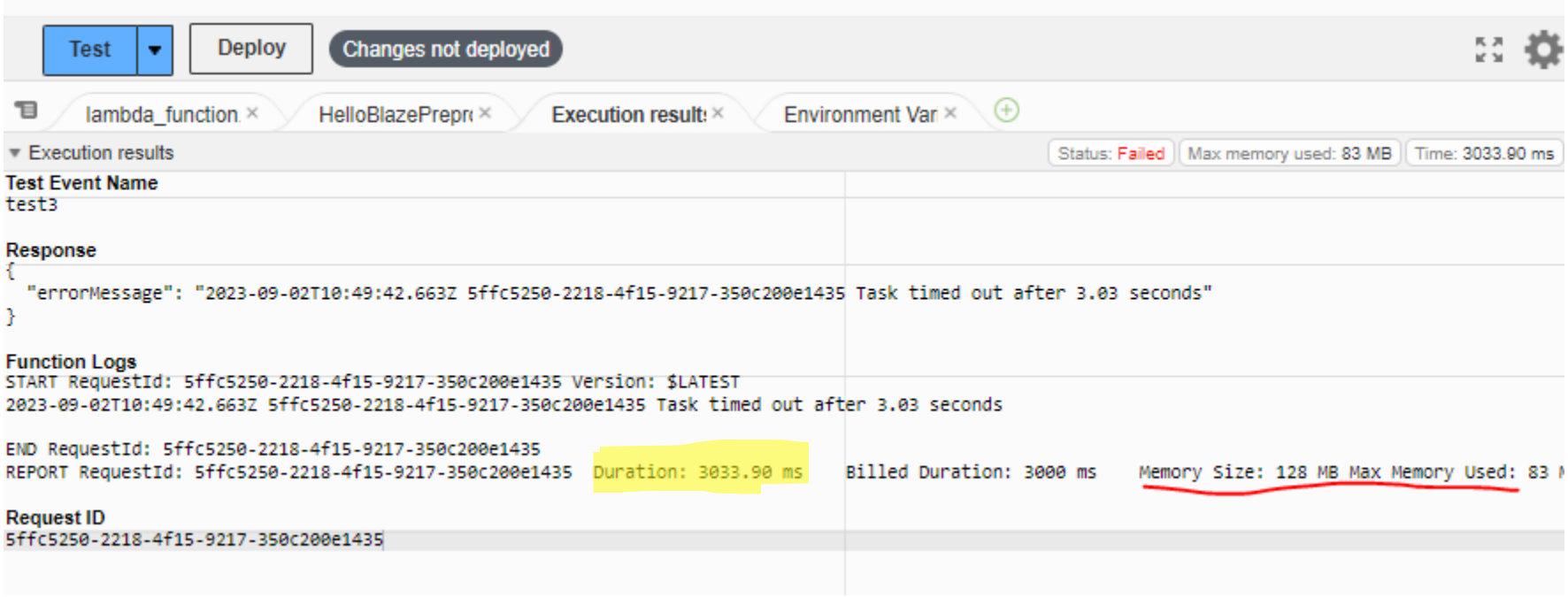
Invoking Lambda function Through a test event (Synchronous Invocation)

Steps:

- Invoke the test event from lamba_handler function.

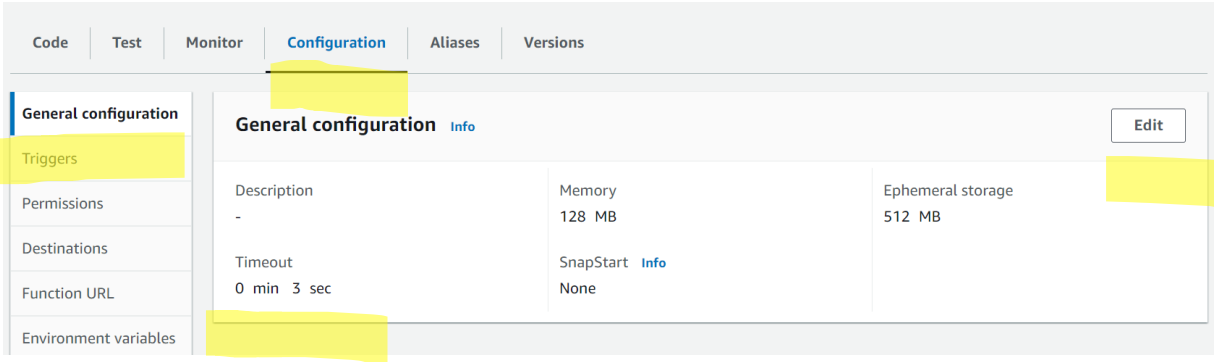
Steps:

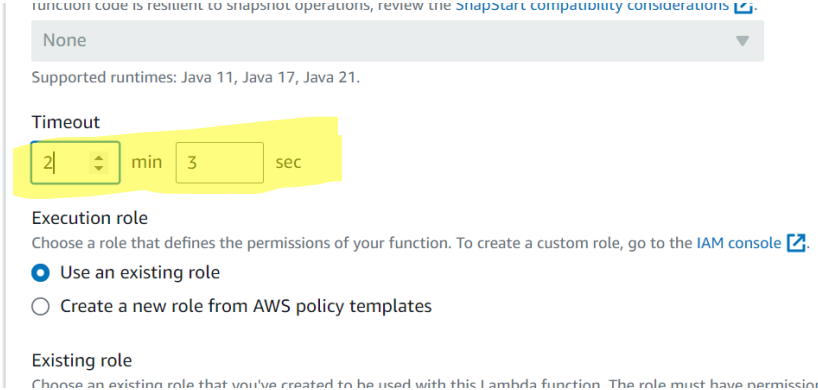
- A problem might appear → Lambda function low execution time



Steps:

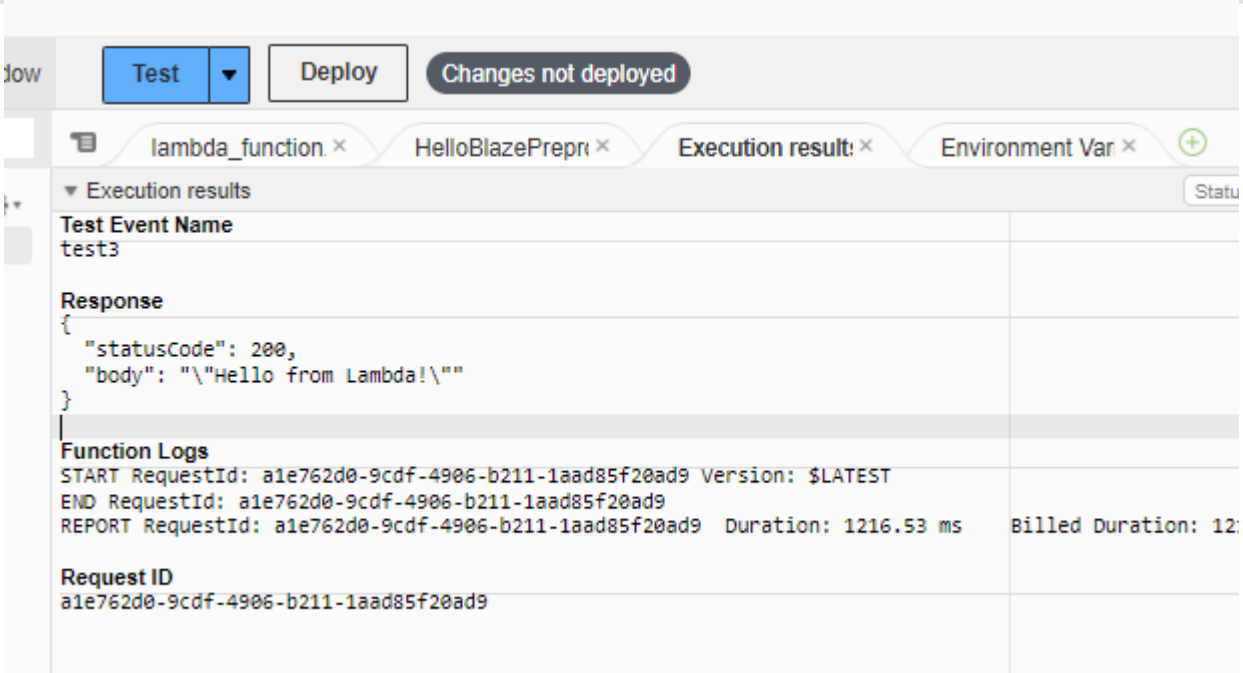
- A solution would be increasing the reserved memory for this lambda function.





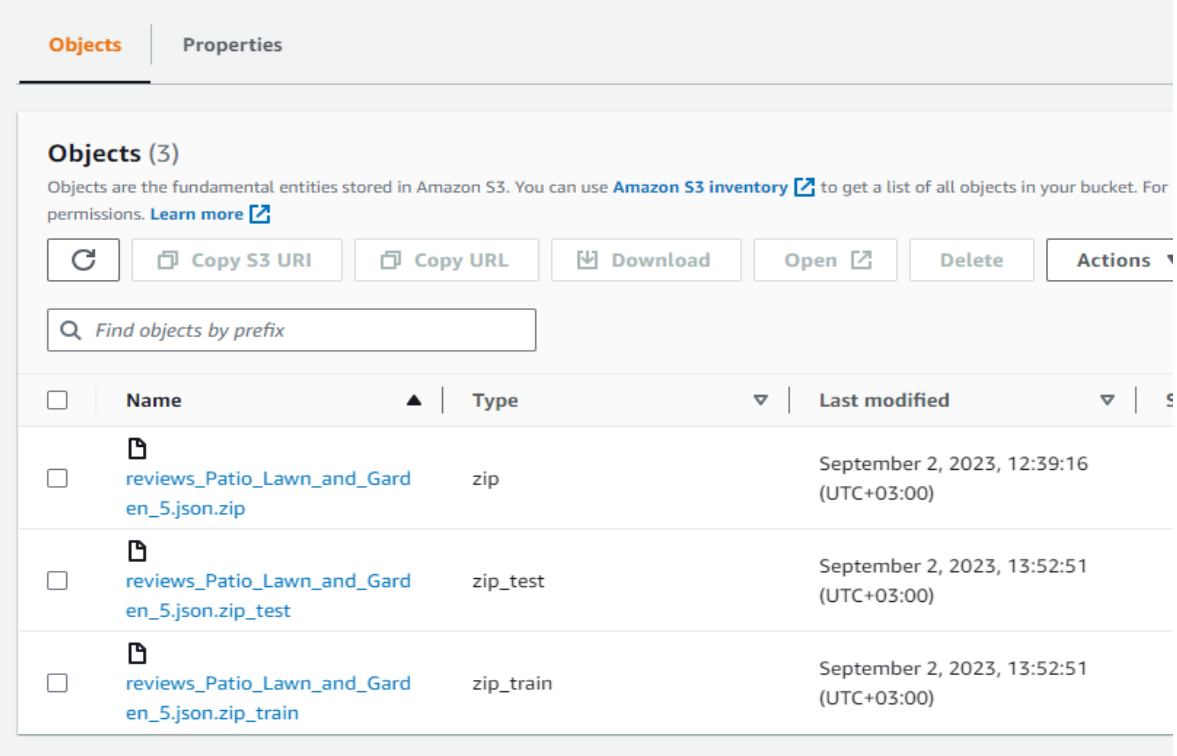
Steps:

- Re-Run test event



Steps:

- Make sure Output is generated



Invoking Lambda function From a Jupyter Notebook (Event invocation) (Synchronous Invocation)

Steps:

- Event Invocation using boto3 SDK

Invoke Lambda_Function

In [13]:
```
import json
payload = {'key':'value'}
payload_bytes =  json.dumps(payload).encode('utf-8')

response = client.invoke(FunctionName = 'Lambda_Function2', #Lambda function name
    InvocationType ='Event',
    Payload = payload_bytes
                         )
```

In [14]:
```
response
```

Out[14]:
```
{'ResponseMetadata': {'RequestId': '0463a5c3-77e3-4b55-afcb-bd36be24d17e',
  'HTTPStatusCode': 202,
  'HTTPHeaders': {'date': 'Sat, 02 Sep 2023 12:38:52 GMT',
   'content-length': '0',
   'connection': 'keep-alive',
   'x-amzn-requestid': '0463a5c3-77e3-4b55-afcb-bd36be24d17e',
   'x-amzn-remapped-content-length': '0',
   'x-amzn-trace-id': 'root=1-64f32cdb-569743a24942981b32636de3;sampled=0'},
  'RetryAttempts': 0},
 'StatusCode': 202,
 'Payload': <botocore.response.StreamingBody at 0x7f1a3e9919f0>}
```

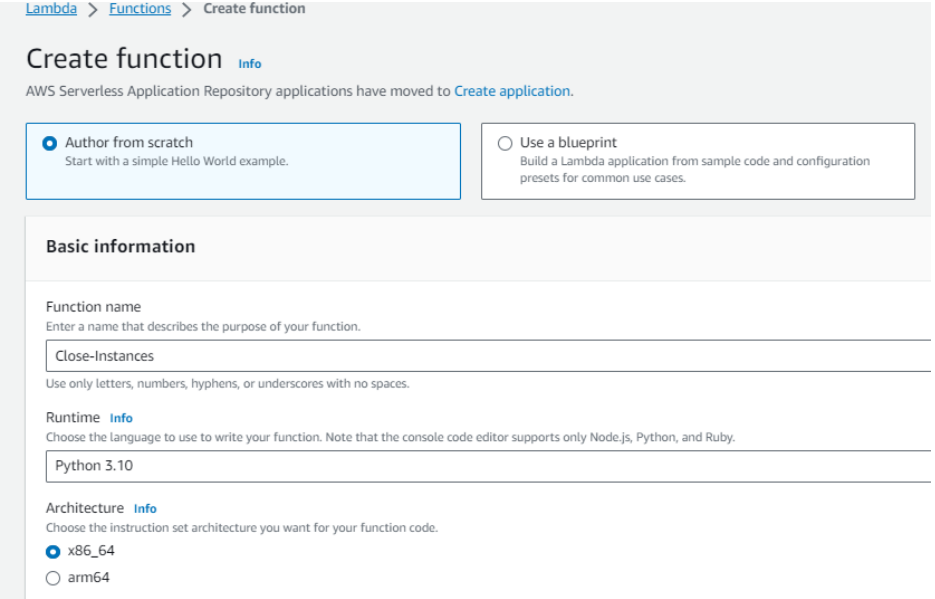
Invoking Lambda function Through a CloudWatch Event Rules (Synchronous Invocation)

Example:

Create a cloud watch event that closes Notebook instances that are open each 10 minutes.

Steps:

Create a Lambda function that closes Notebooks



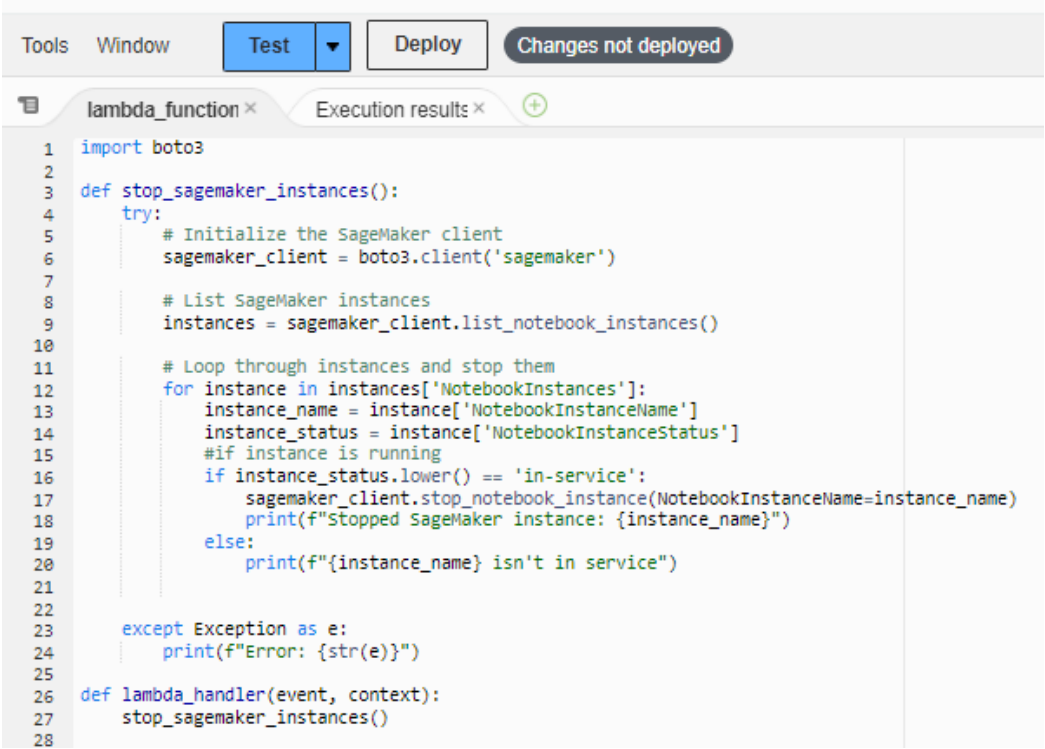
Steps:

Add the following code to lambda_function.py

```
import boto3
def stop_sagemaker_instances():
        try:
                # Initialize the SageMaker client
                sagemaker_client = boto3.client('sagemaker')

                #Replace 'your-region' with your AWS region # List SageMaker instances
                instances = sagemaker_client.list_notebook_instances()

                 # Loop through instances and stop them
                 for instance in instances['NotebookInstances']:
                        instance_name = instance['NotebookInstanceName']
                        instance_status = instance['NotebookInstanceStatus']

                        if instance_status.lower() == 'in-service':
                                sagemaker_client.stop_notebook_instance(
                                        NotebookInstanceName=instance_name)
                                print(f"Stopped SageMaker instance: {instance_name}")
                        else:
                                print(f" {instance_name} isn't in service")
        except Exception as e:
                print(f"Error: {str(e)}")
def lambda_handler(event, context):
        stop_sagemaker_instances()
```

Steps:

Add the following code to lambda_function.py

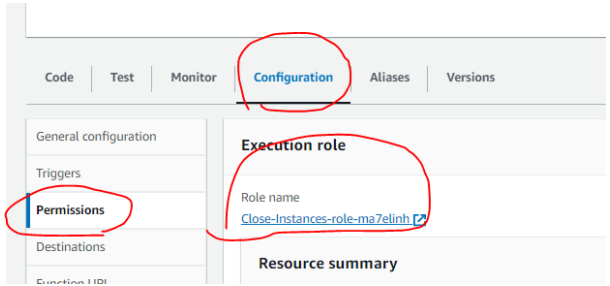


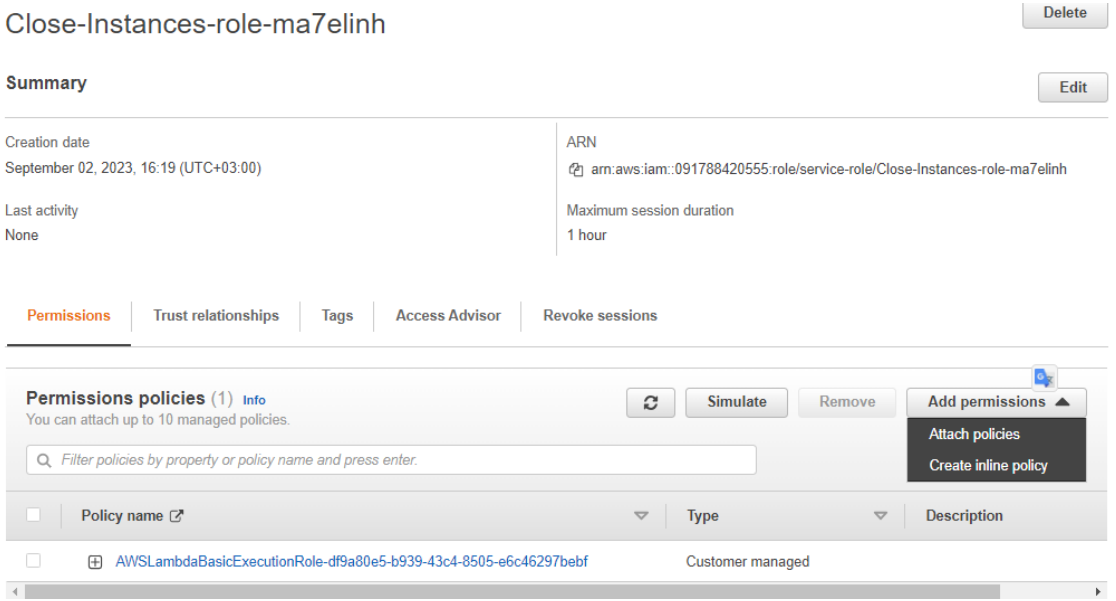

```python
import boto3

def stop_sagemaker_instances():
    try:
        # Initialize the SageMaker client
        sagemaker_client = boto3.client('sagemaker')

        # List SageMaker instances
        instances = sagemaker_client.list_notebook_instances()

        # Loop through instances and stop them
        for instance in instances['NotebookInstances']:
            instance_name = instance['NotebookInstanceName']
            instance_status = instance['NotebookInstanceStatus']
            #if instance is running
            if instance_status.lower() == 'in-service':
                sagemaker_client.stop_notebook_instance(NotebookInstanceName=instance_name)
                print(f"Stopped SageMaker instance: {instance_name}")
            else:
                print(f"{instance_name} isn't in service")


    except Exception as e:
        print(f"Error: {str(e)}")

def lambda_handler(event, context):
    stop_sagemaker_instances()
```

Steps:

Give the right permissions to the lambda function corresponding role.

- **AmazonSageMakerFullAccess** gives permission to notebooks and deleting them

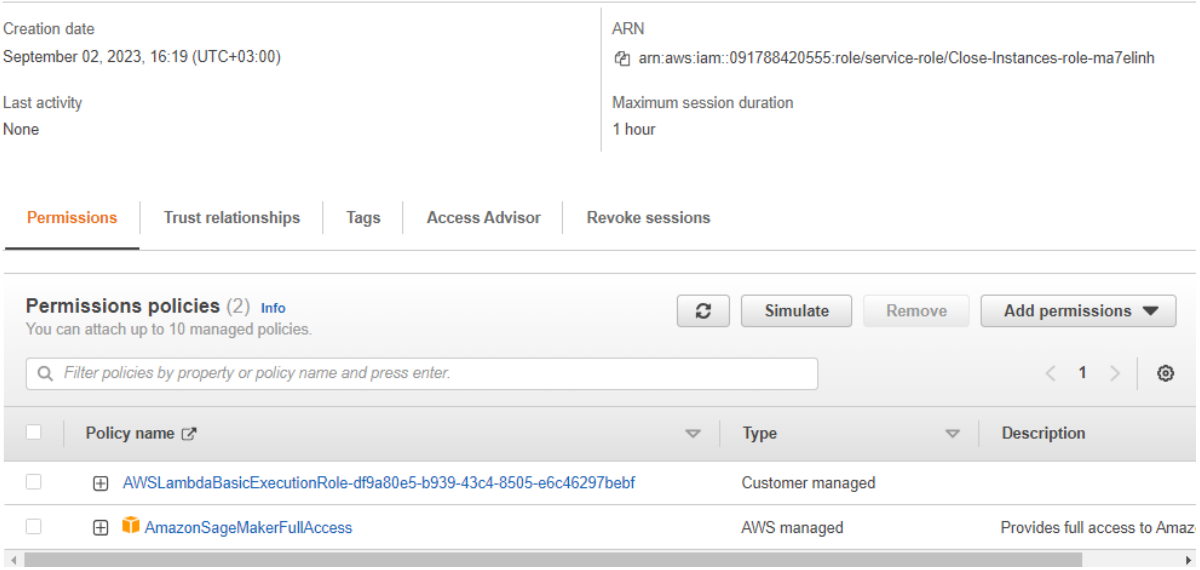




Steps:

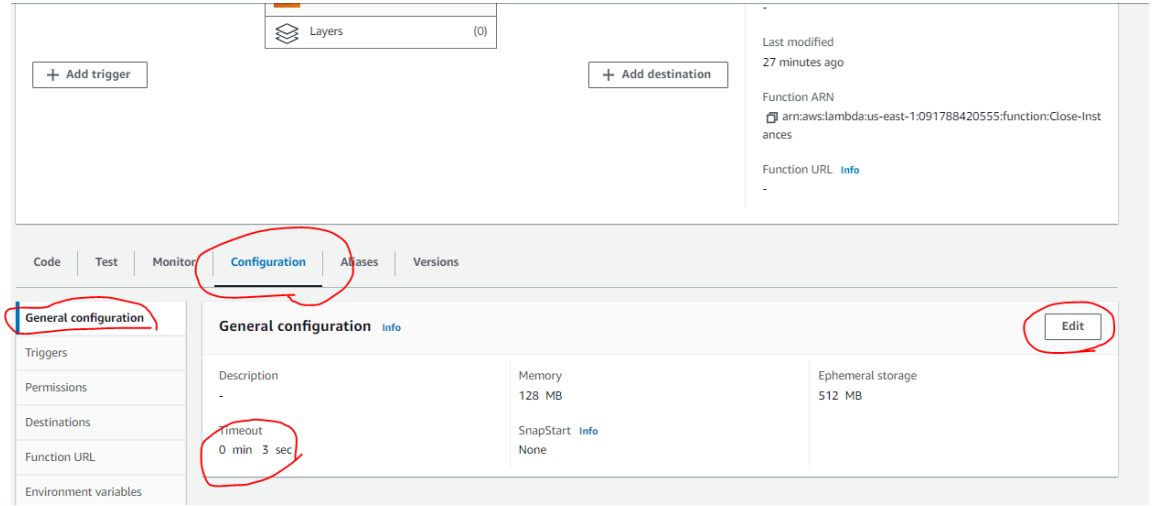
Give the right permissions to the lambda function corresponding role.

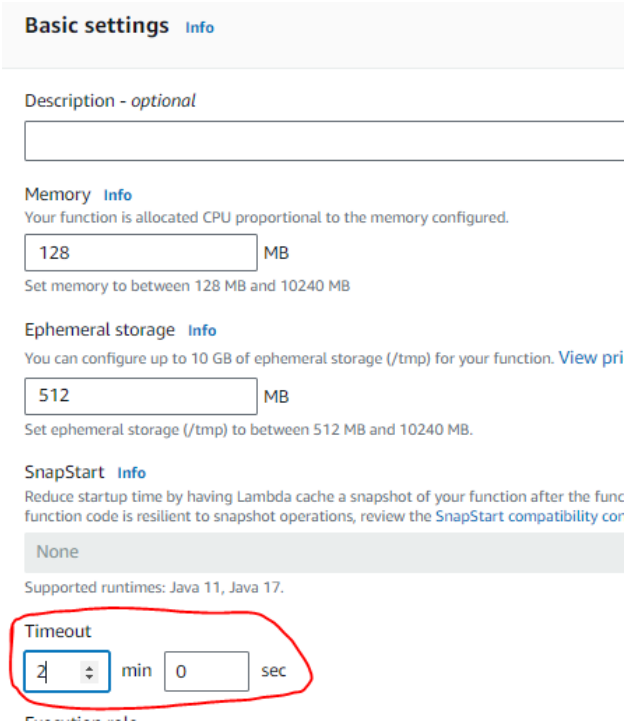
- **AmazonSageMakerFullAccess** gives permission to notebooks and deleting them



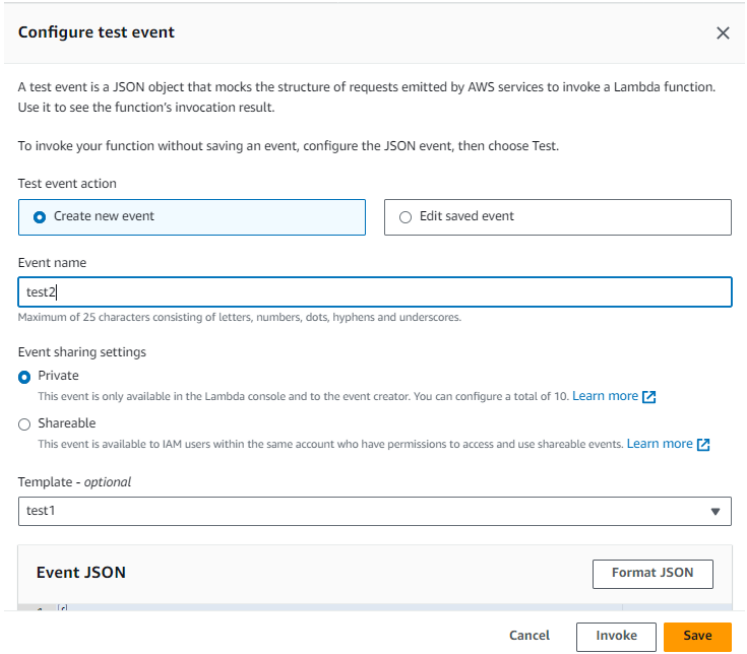
Steps:

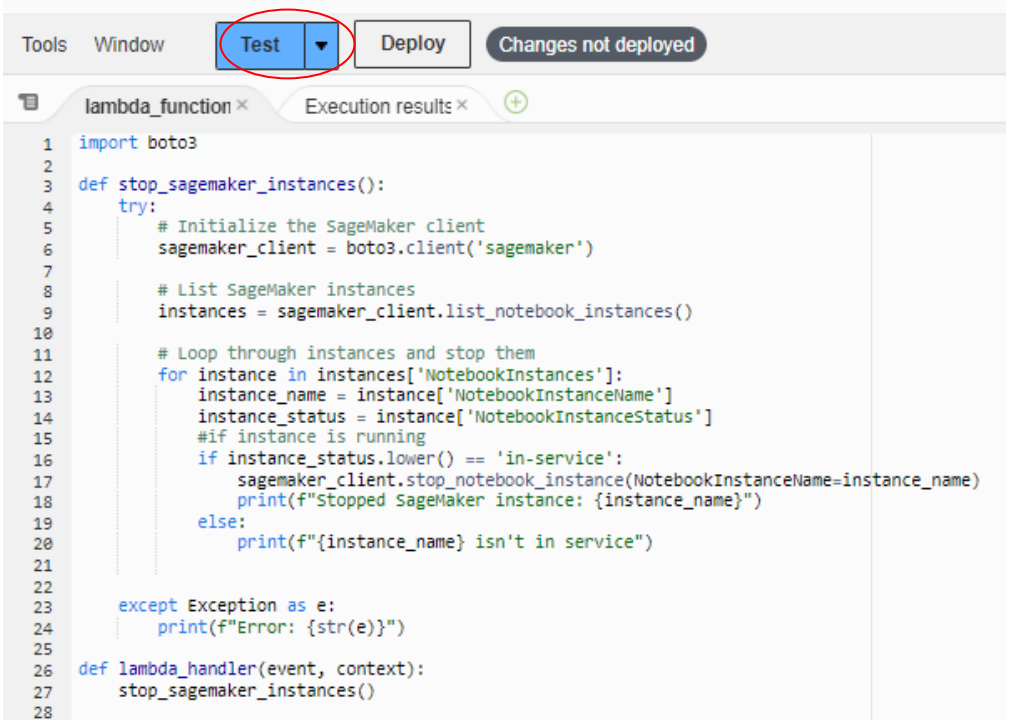
Give lambda more time to execute maybe 2 mins is enough.



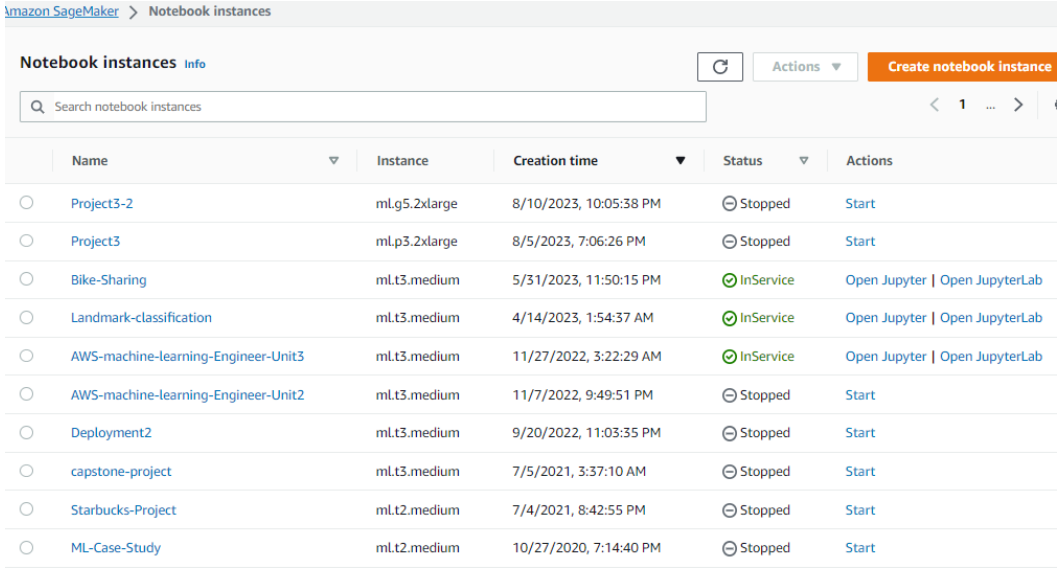


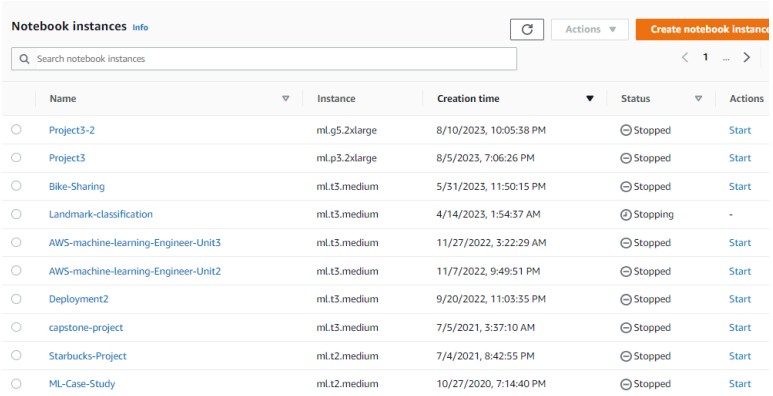
You can test the event now with the test event as shown. It will close all instances without scheduling. So **be careful**



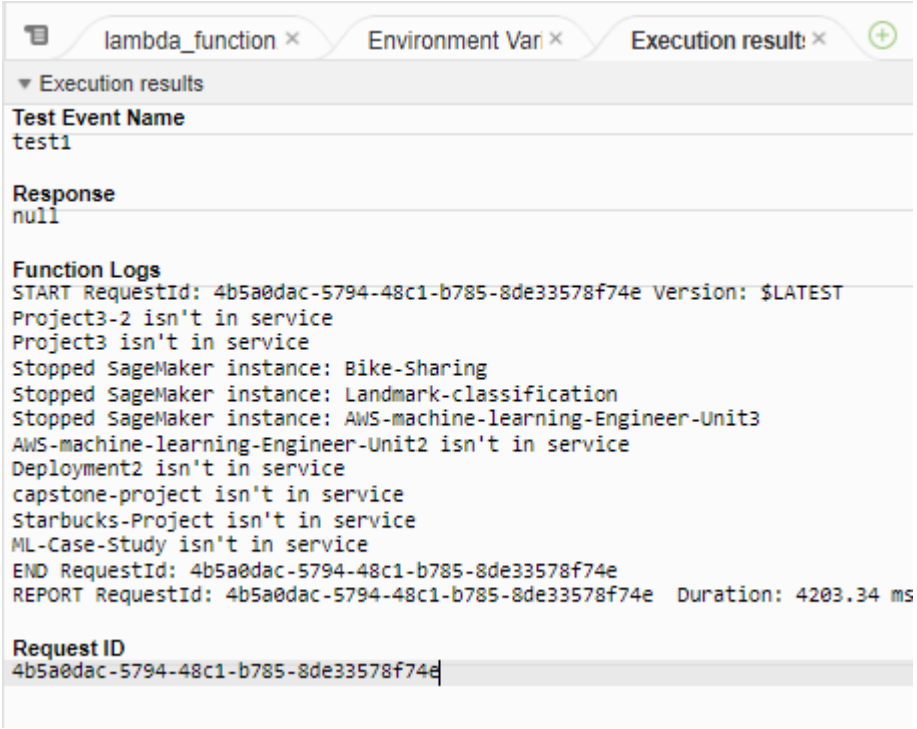


Before & After



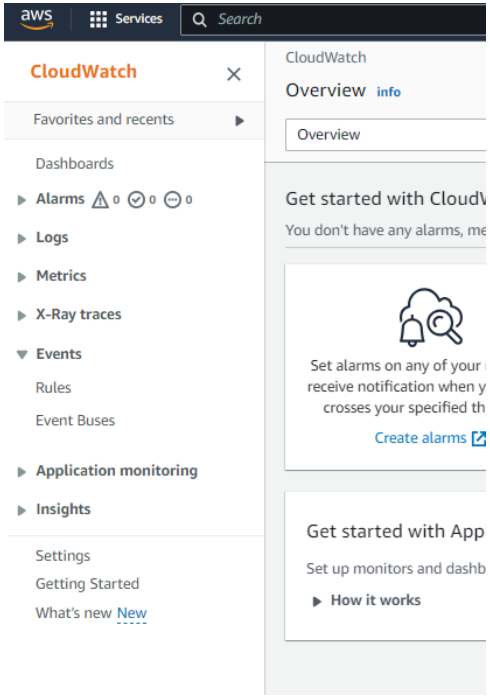


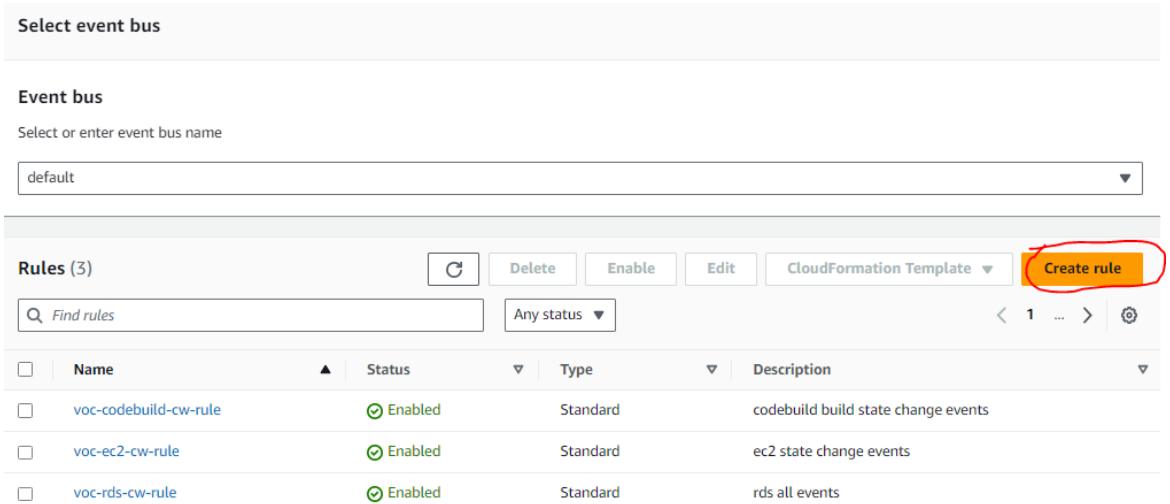
Code output of lambda function



Steps:

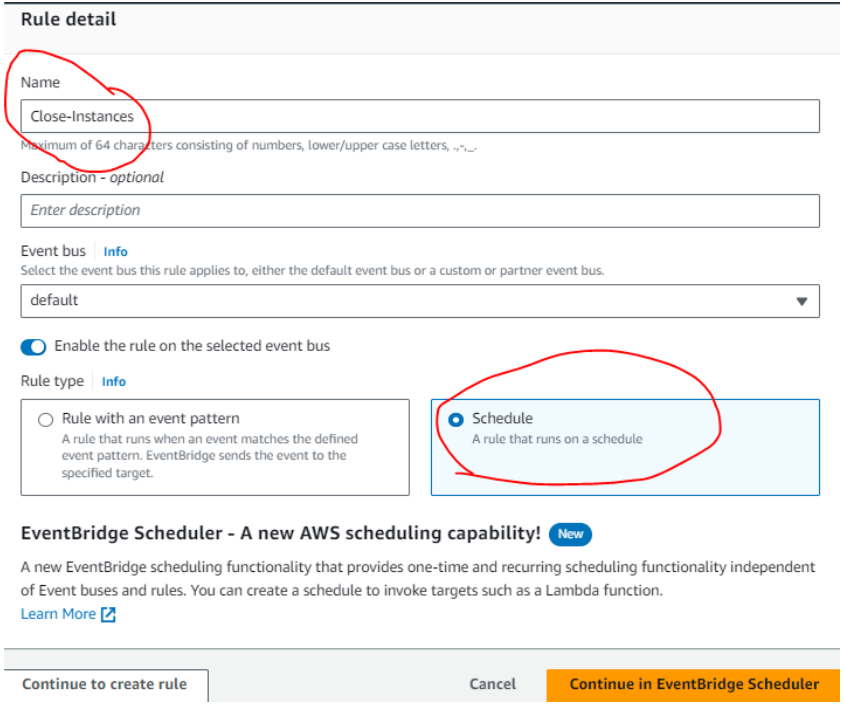
Create a Cloudwatch scheduled event to trigger/invoke the LambdaFunction

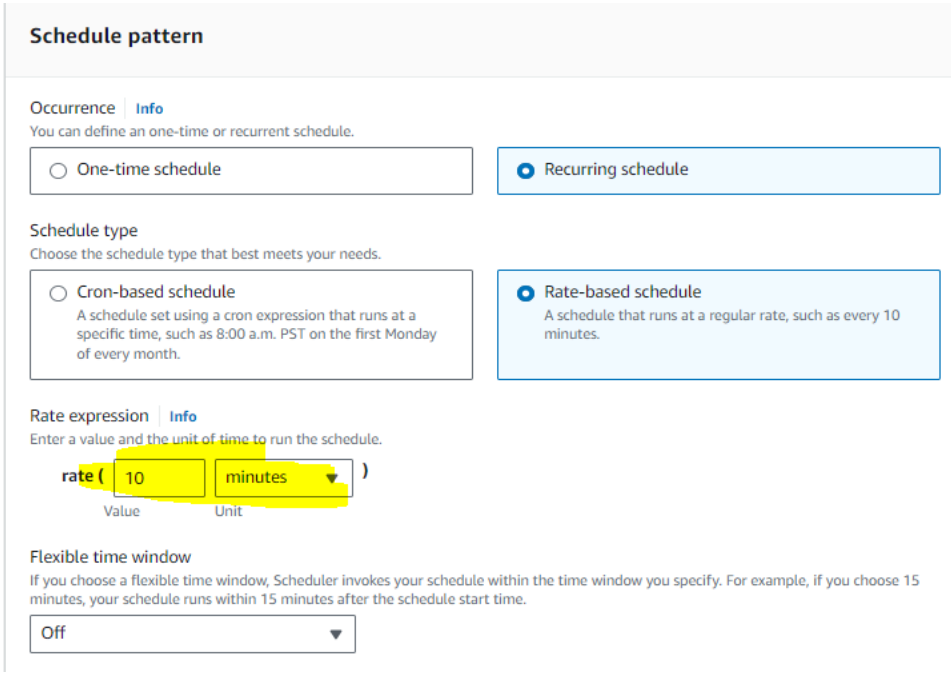




Steps:

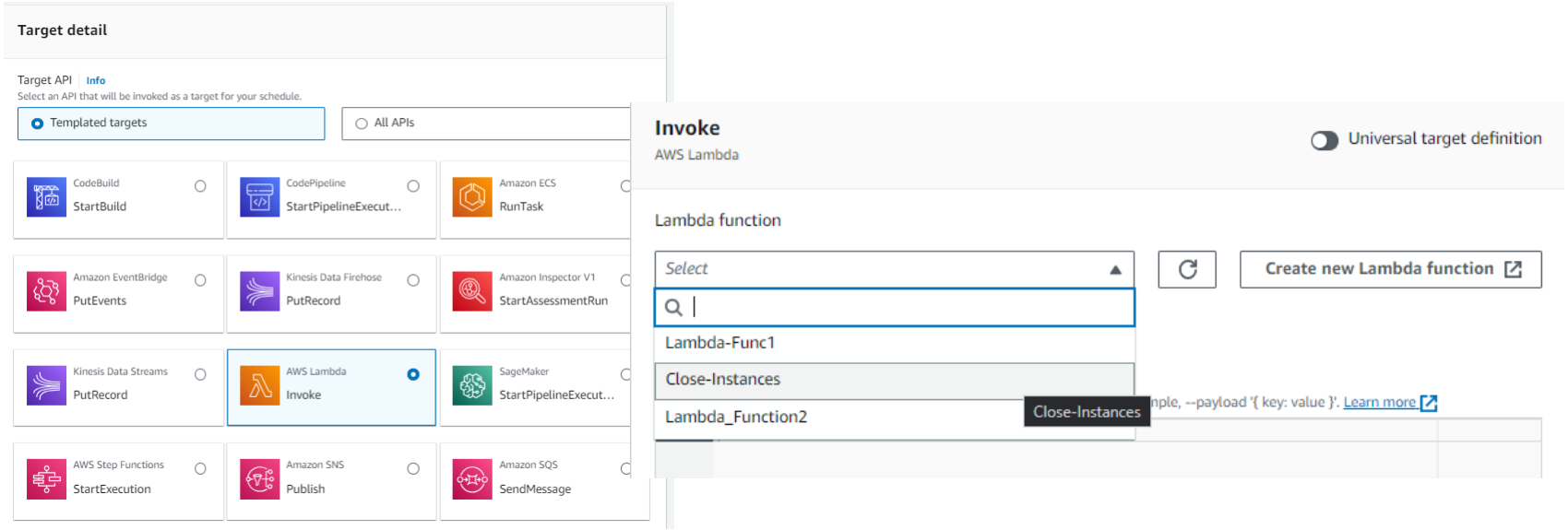
Create a Cloudwatch scheduled event to trigger/invoke the LambdaFunction





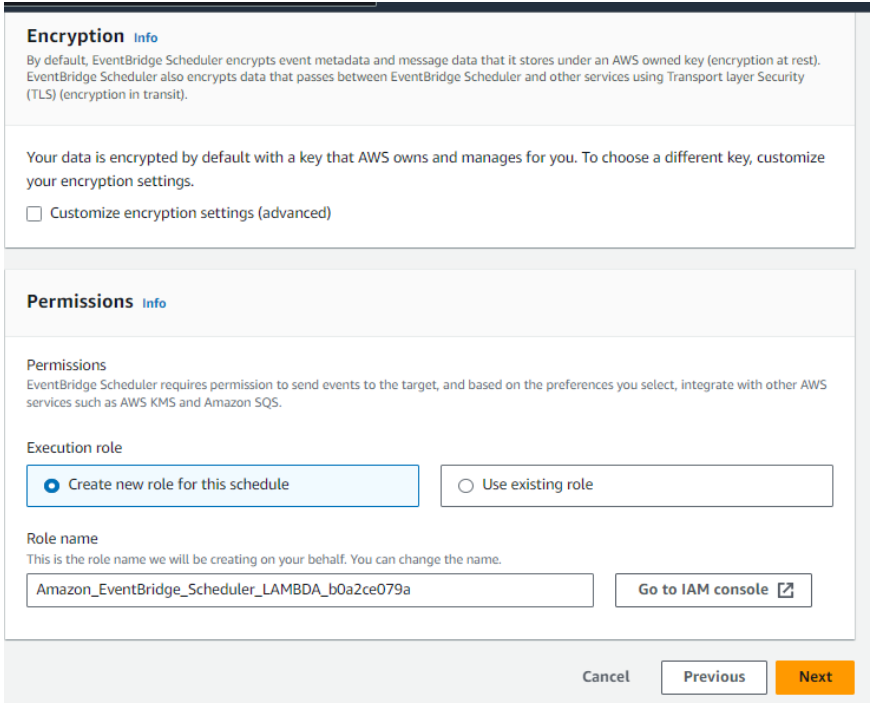
Steps:

Create a Cloudwatch scheduled event to trigger/invoke the LambdaFunction



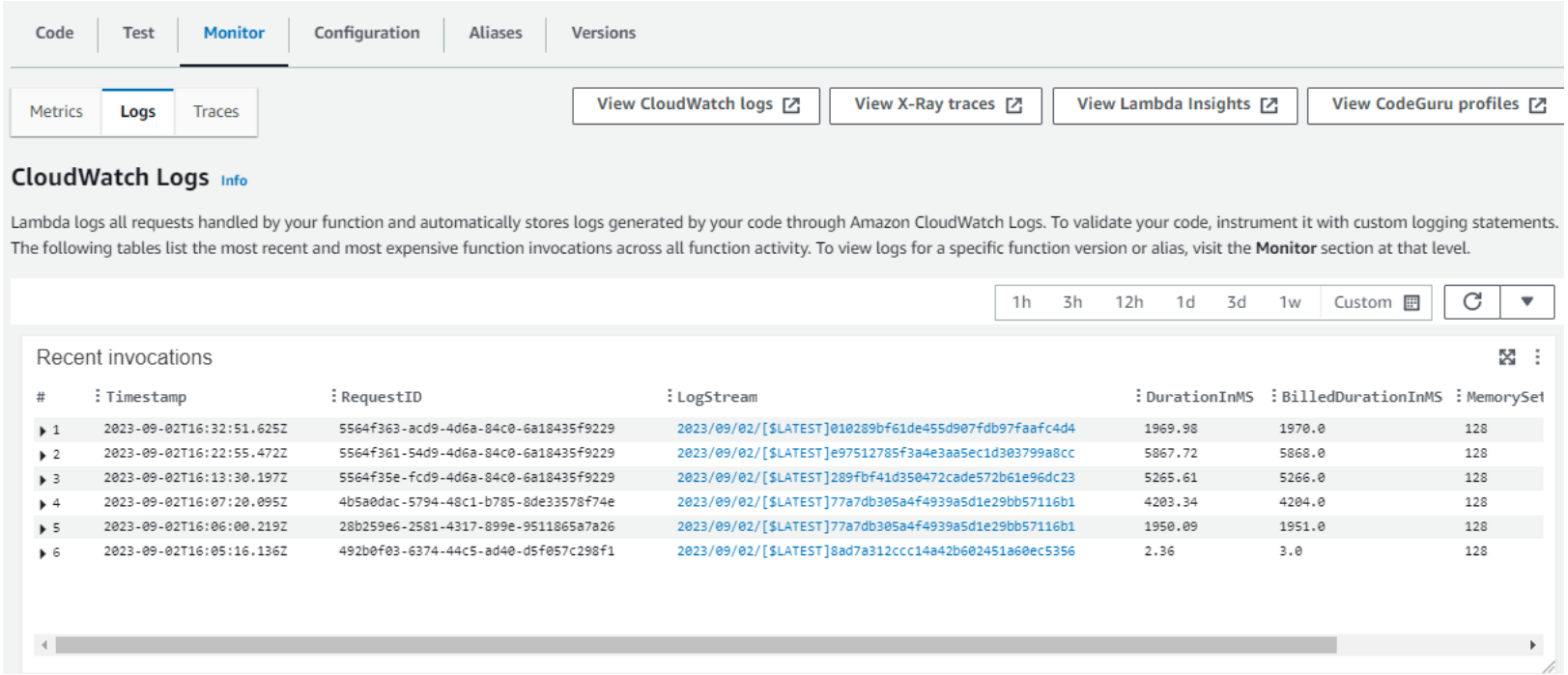
Note:

This task won't be implemented using the given AWS accounts. Use a personal account to have the authority to create a cloud watch event



Steps:

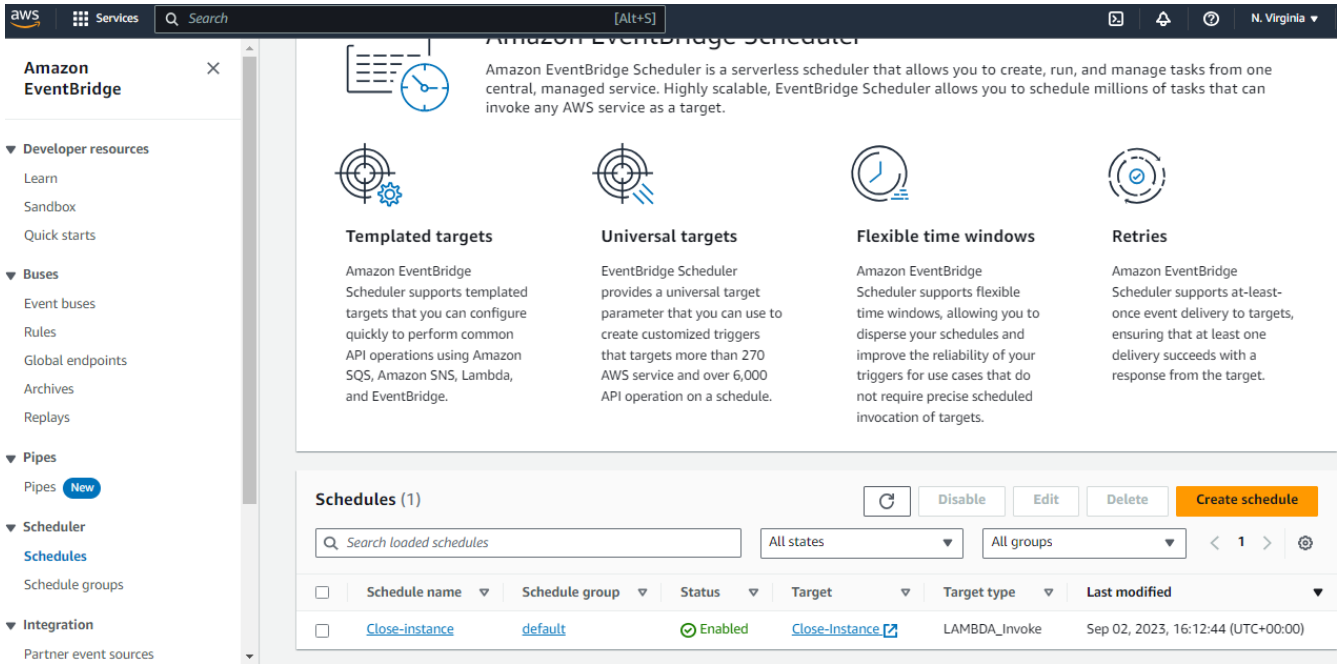
Monitor your lambda function after a CloudWatch Event is created



#	Timestamp	RequestID	LogStream	DurationInMS	BilledDurationInMS	MemorySet
1	2023-09-02T16:32:51.625Z	5564f363-acd9-4d6a-84c0-6a18435f9229	2023/09/02/[$LATEST]010289bf61de455d907fdb97faafc4d4	1969.98	1970.0	128
2	2023-09-02T16:22:55.472Z	5564f361-54d9-4d6a-84c0-6a18435f9229	2023/09/02/[$LATEST]e97512785f3a4e3aa5ec1d303799a8cc	5867.72	5868.0	128
3	2023-09-02T16:13:30.197Z	5564f35e-fcd9-4d6a-84c0-6a18435f9229	2023/09/02/[$LATEST]289fbf41d350472cade572b61e96dc23	5265.61	5266.0	128
4	2023-09-02T16:07:20.095Z	4b5a0dac-5794-48c1-b785-8de33578f74e	2023/09/02/[$LATEST]77a7db305a4f4939a5d1e29bb57116b1	4203.34	4204.0	128
5	2023-09-02T16:06:00.219Z	28b259e6-2581-4317-899e-9511865a7a26	2023/09/02/[$LATEST]77a7db305a4f4939a5d1e29bb57116b1	1950.09	1951.0	128
6	2023-09-02T16:05:16.136Z	492b0f03-6374-44c5-ad40-d5f057c298f1	2023/09/02/[$LATEST]8ad7a312ccc14a42b602451a60ec5356	2.36	3.0	128

Steps:

Don’t forget to terminate your scheduler



Break (10 minutes)

Satisfaction Survey

Step Functions

State Machine:

Workflow/ Repeated pattern of activity (Ex: Pattern Of activation Functions)

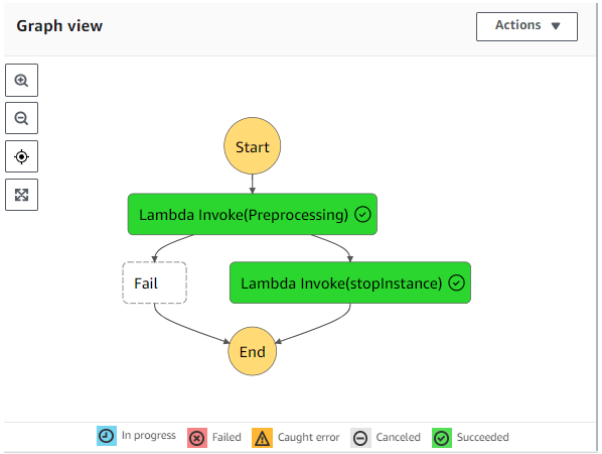
Task:

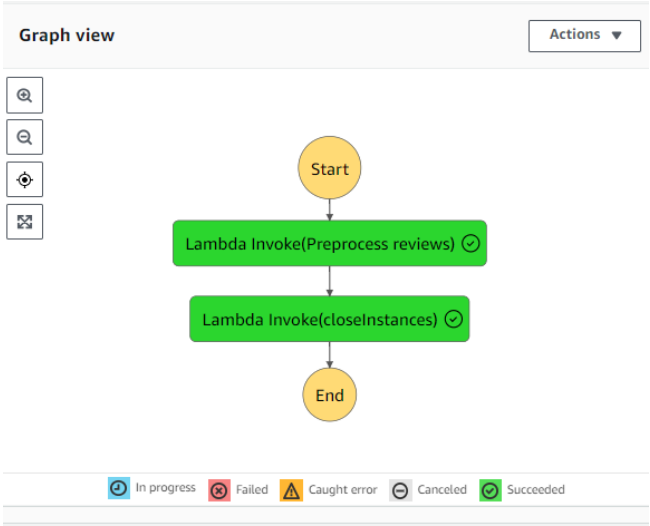
Single unit / activity in the workflow.

Step Function achieves:

1- Orchestration: Ordered execution of activities

2- Branching: Determination of which path to take based on the prior task's state





Problems With Step Functions

Needs Knowledge of Amazon state Language
Expensive when executed
Not Compatible with other similar orchestration tools

When to consider?

Complex processes to be implemented less frequently

When to consider other tools?

Simple processes to be implemented more frequently

Task: Create a step function that preprocess Toys' Reviews and after that closes all SageMaker Instances (Console & SDK)

ML workflow

Exercise: Create a workflow to preprocess the Breast Cancer dataset and to run an estimator for training an XGBOOST model

Any Question?

Thank you

UDACITY