Hello World using AWS Lambda:

- Go to AWS Console and use the Lambda template.
- Deploy and Test the Hello world function and make sure it is successful.

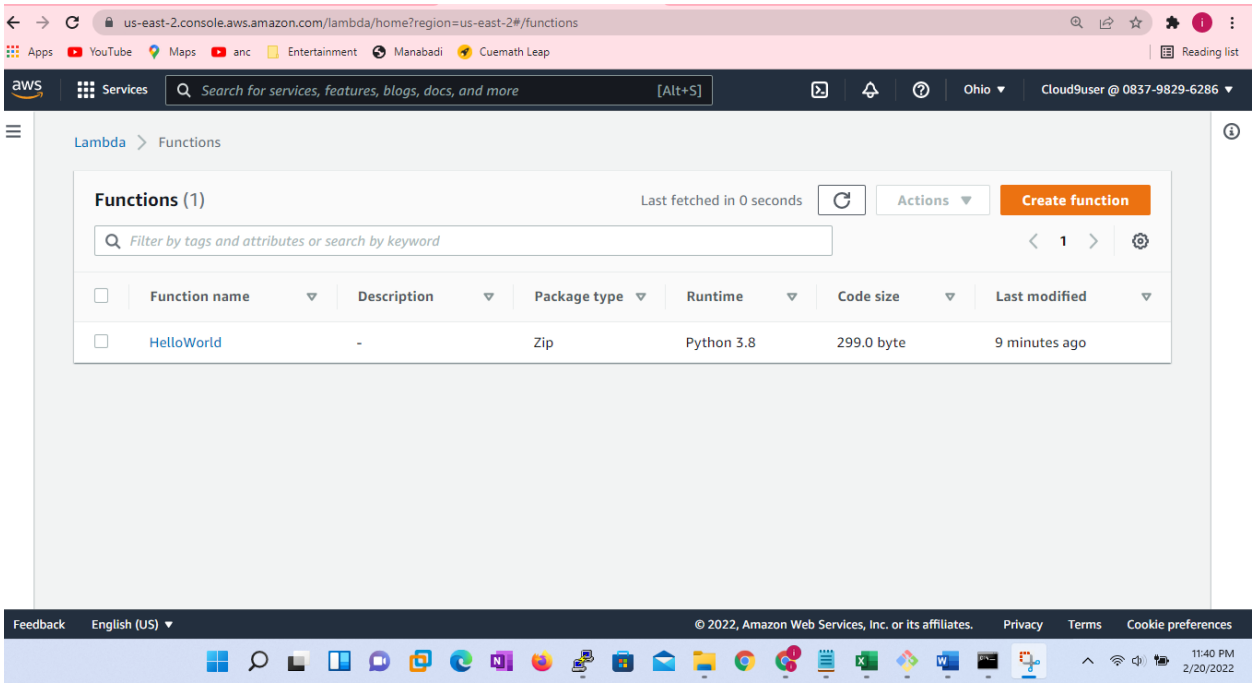


Figure 1Create the Hello world function

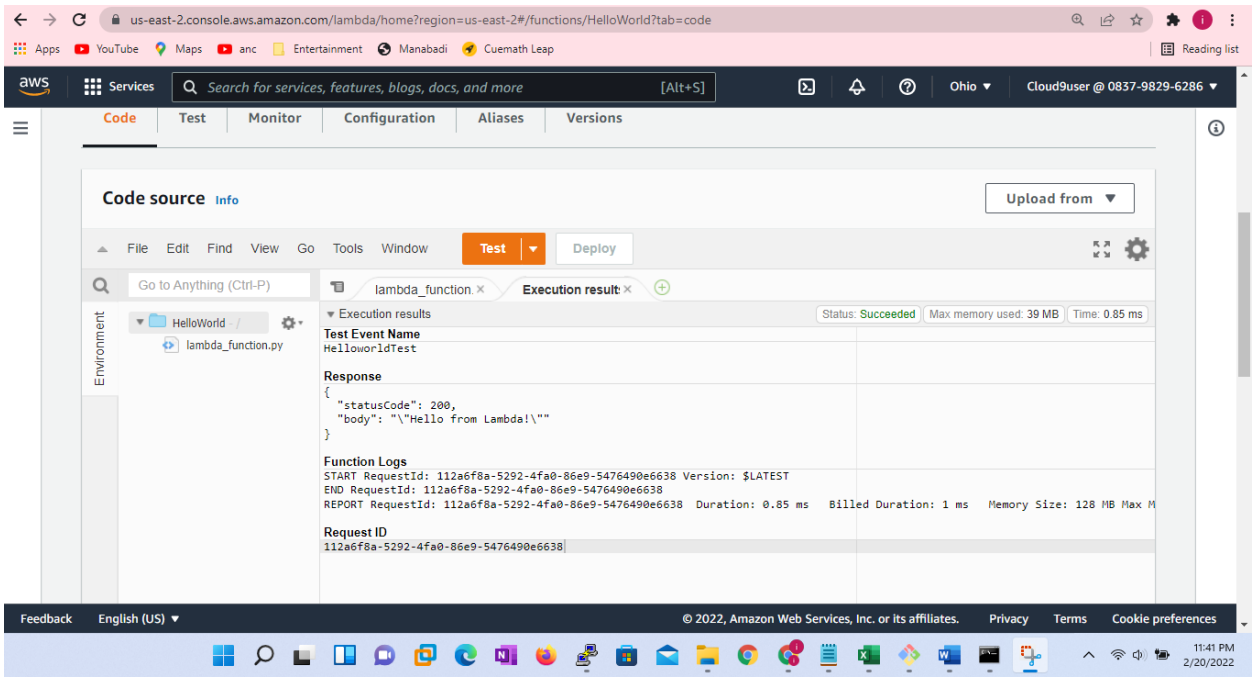


Figure 2Tested and deployed the Hello World function

Setup Project for local development

Create folder for the project - **ghactivity-downloader**
Create virtual environment for this project - **ghad-venv**

python3 -m venv ghad-venv
source ghad-venv/bin/activate

We will install boto3 in the default location within the virtual environment. We need not include it as part of the bundle that will be deployed as a lambda function.

pip install boto3
pip install requests

We will install requests as part of **lambdalib** folder. It will be included as part of the bundle that will be deployed as a lambda function.

mkdir Ghalib
pip install requests -t Ghalib

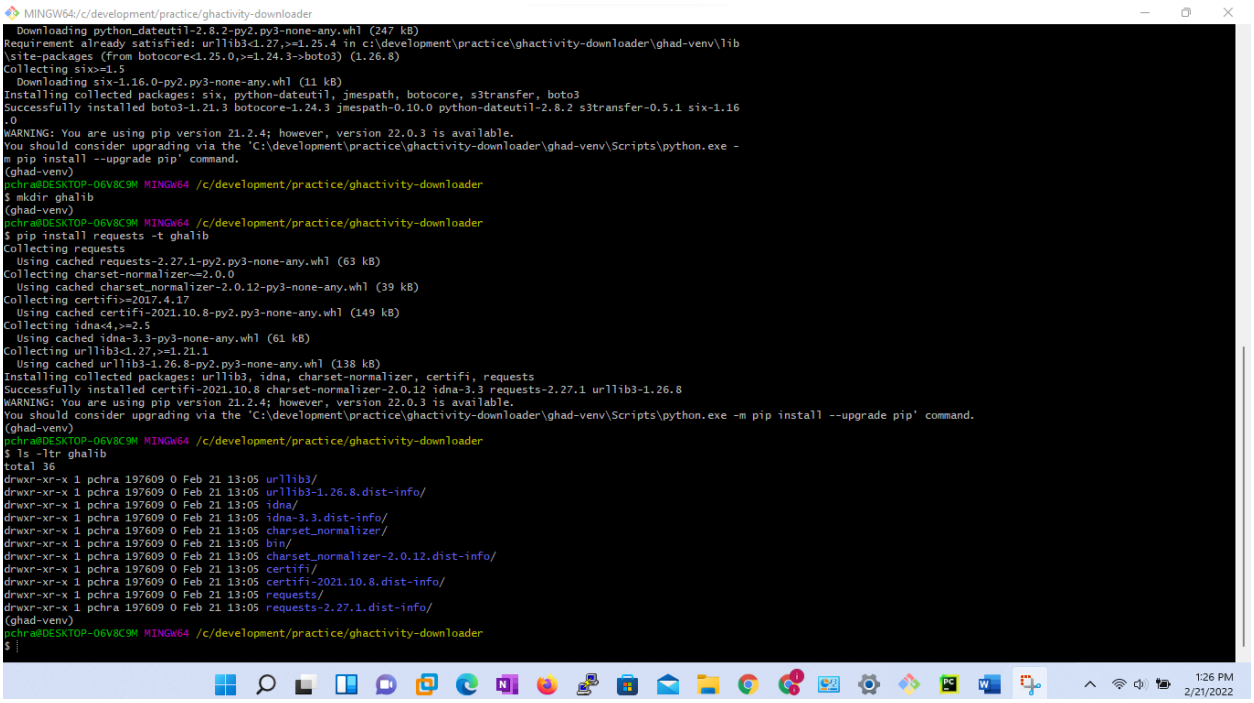


Figure 3Environment is created

Deploy Project to AWS Lambda console:

we can deploy the locally developed Lambda Function to AWS Lambda Web Console.

- You need to build the zip file with the source code.

```
zip -r ghactivity-downloader.zip lambda_function.py
```

- Use AWS Lambda Web Console to upload the Zip file.

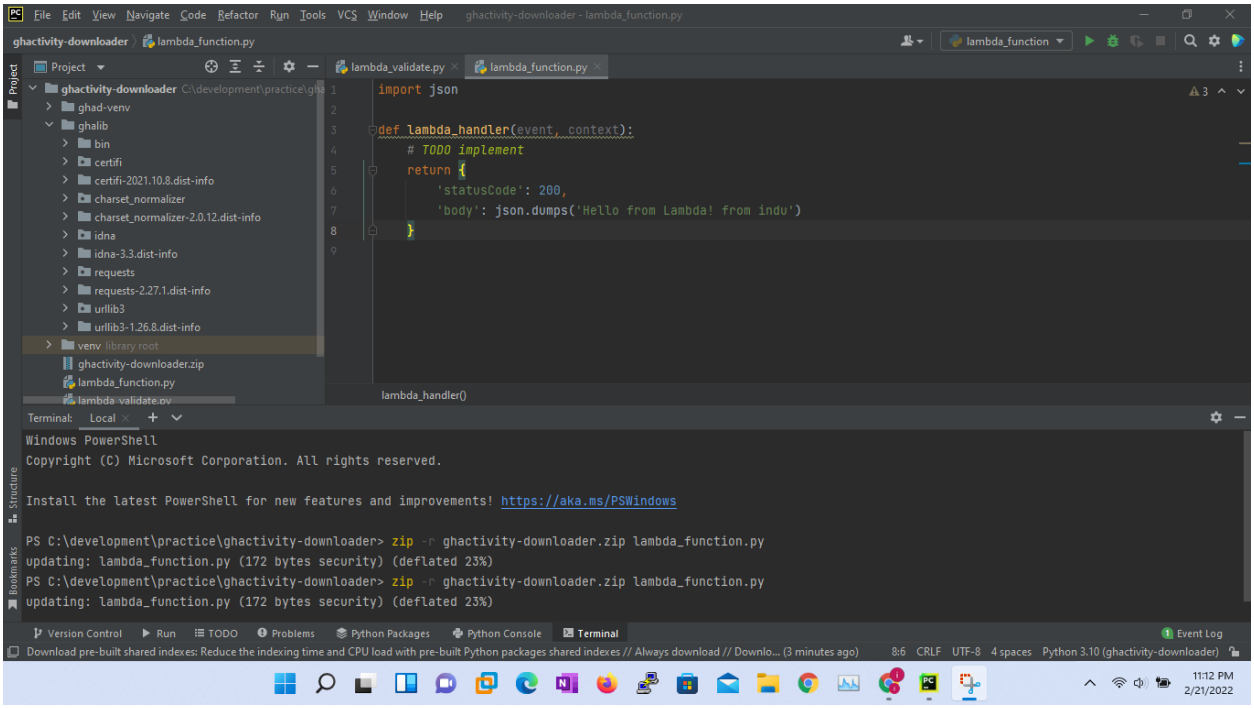
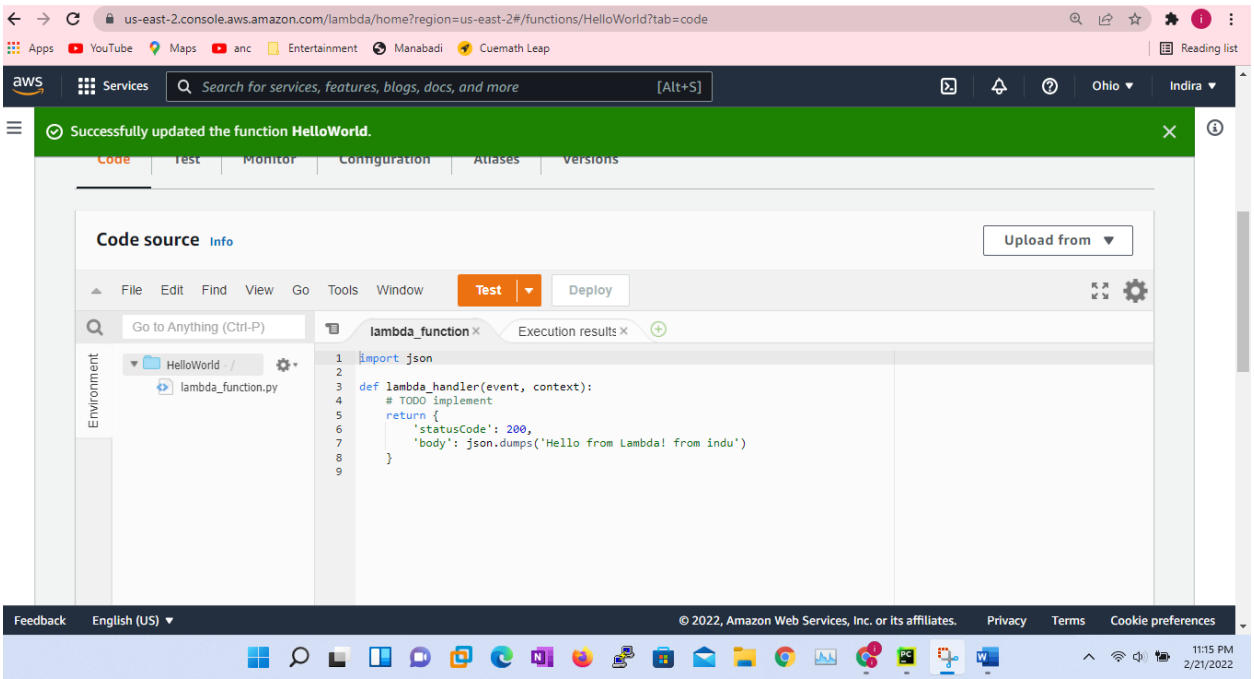


Figure 4Simple Welcome program wrote in pycharm

Uploaded zip code into the aws lambda web console



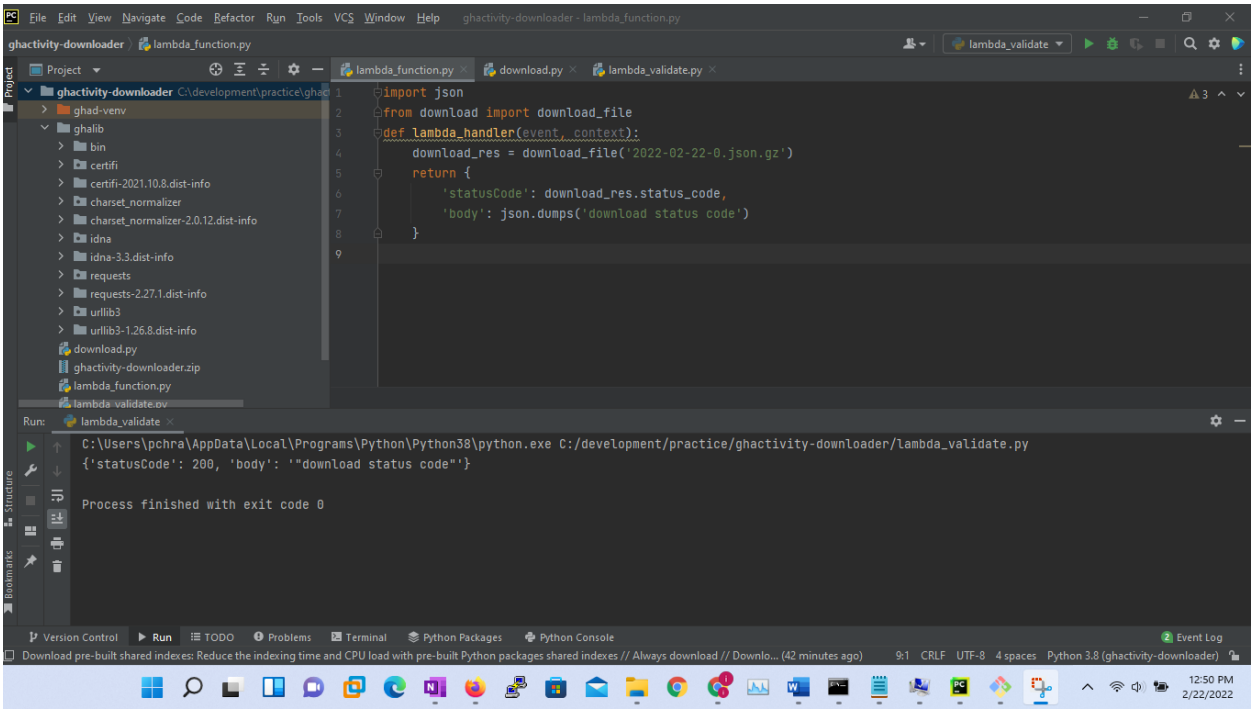
Develop download functionality using requests

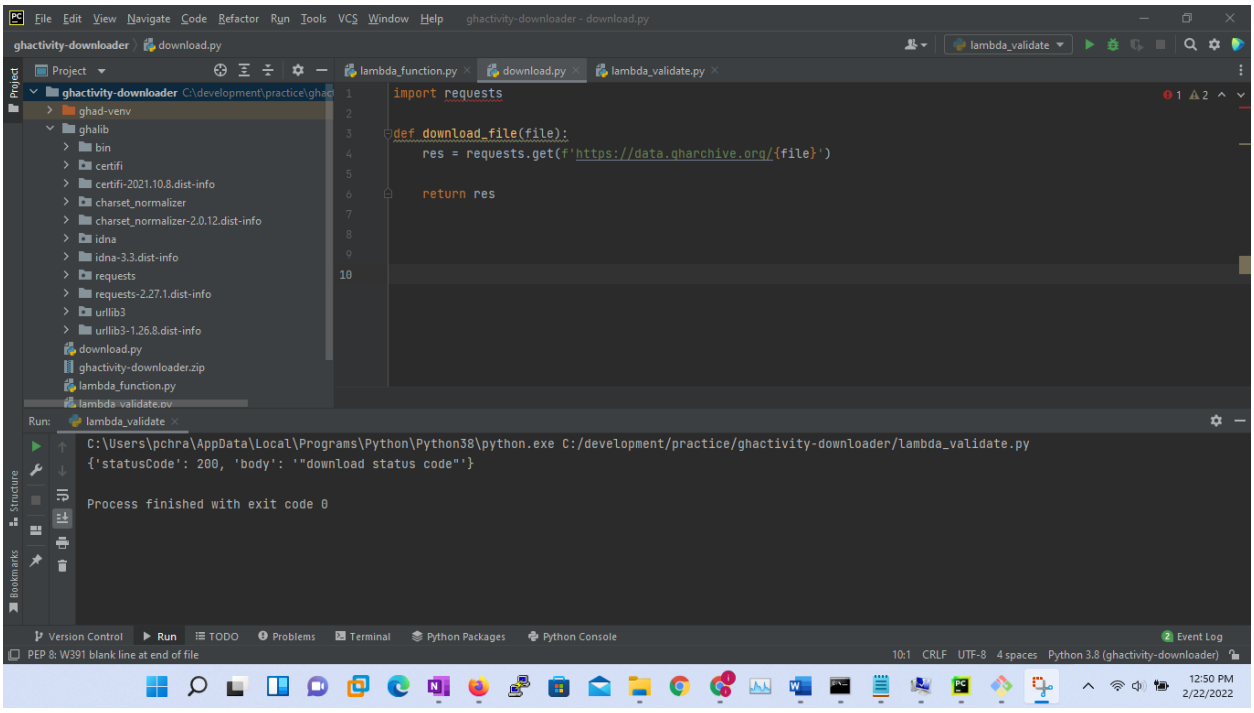
Develop the base functionality to download the zip file using requests library. I have created a new program called download.py for this.

```
import requests
def download_file(file):
res = requests.get(f'https://data.gharchive.org/{file}')
return res
```

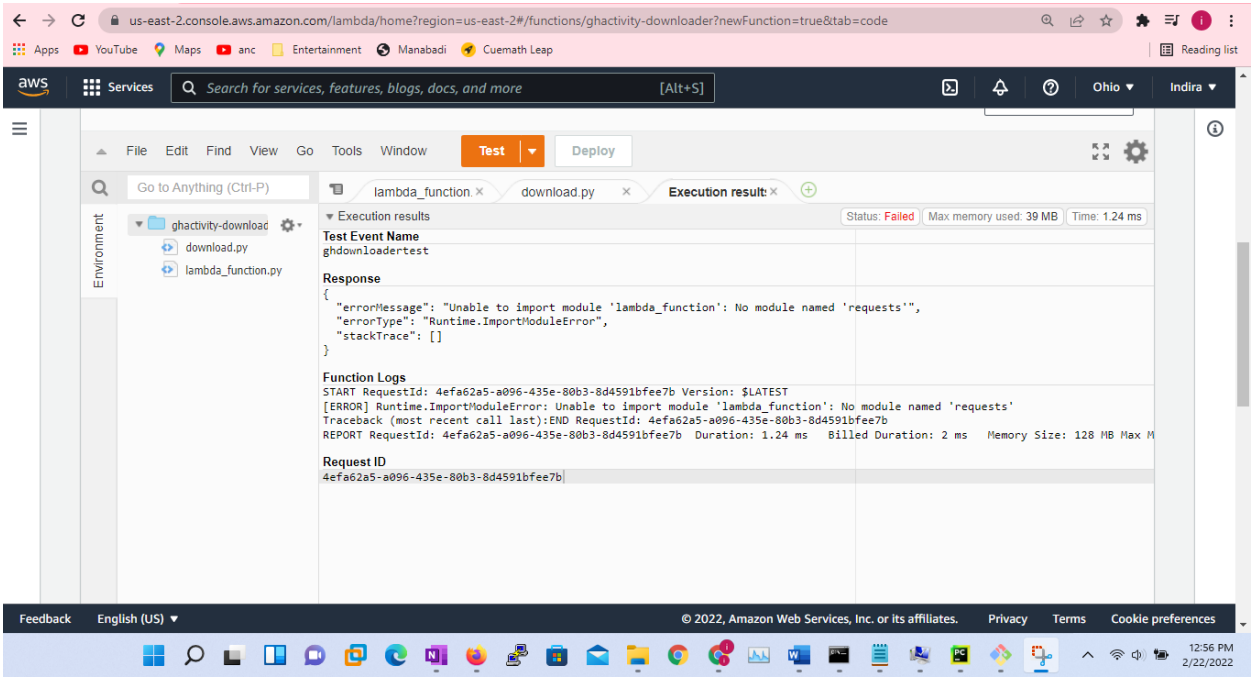
Refactor the code as part of lambda_function.py to invoke the new function and also to capture the response.

```
import json
from download import download_file

def lambda_handler(event, context):

  download_res = download_file('2021-01-29-0.json.gz')

  return {

    'statusCode': download_res.status_code,

    'body': json.dumps('Download status code')

  }
```





Python program is zipped and that zip file is uploaded into the aws lambda it's not failing "unable to import module requests



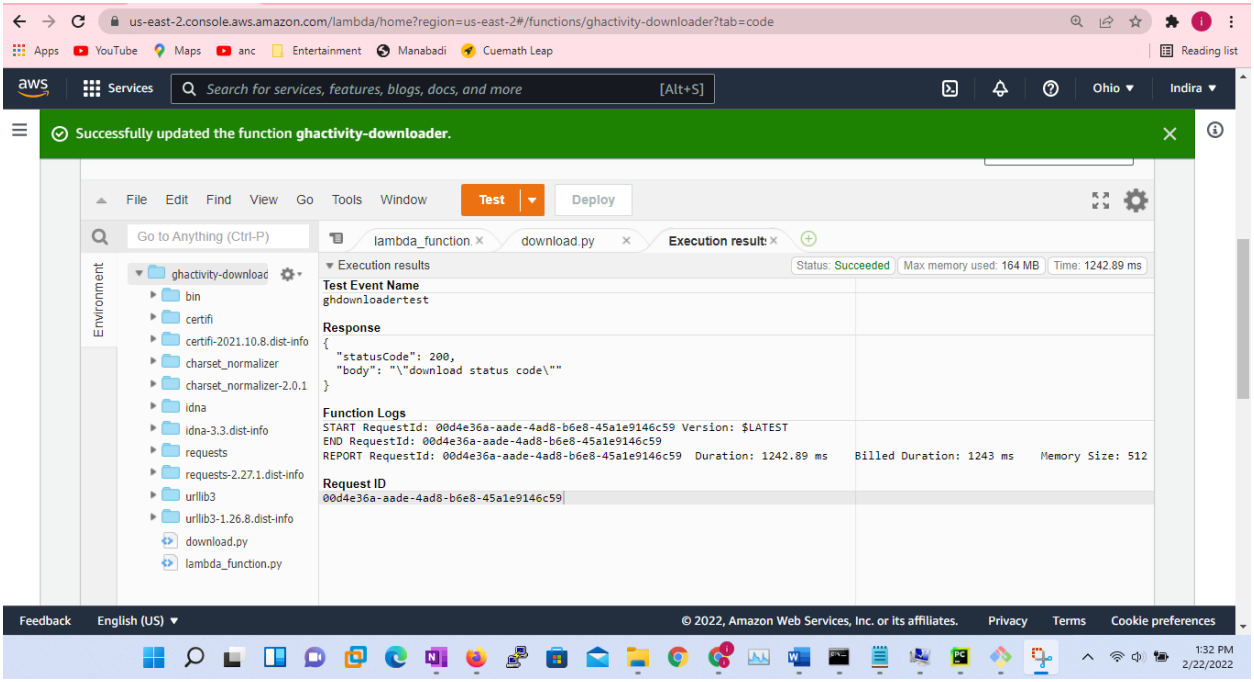
Now to have to include request as part of the zip file then we have to reupload the zip file.

Using 3rd party libraries in AWS Lambda:

We need to go to the folder to build the zip file. Make sure the zip file is created in the base directory of the project and update the zip file with source code.

```
rm ghactivity-downloader.zip
cd Ghalib
zip -r ../ghactivity-downloader.zip .
cd ..
zip -g ghactivity-downloader.zip lambda_function.py download.py
```

- We can upload the zip file to AWS Lambda console and validate successfully. Make sure to increase memory size to 512 MB as demonstrated.



Validating s3 access for local development:

Let us validate s3 access for local development. We need to have the appropriate credentials to access s3 bucket from the local development environment.

- Develop the code to upload the contents of the zip file from GitHub archive to s3. We will create a new file called upload.py.

```
import os
import boto3
import requests
os.environ.setdefault('AWS_DEFAULT', 'itvgithub')
s3_client = boto3.client('s3')
file = '2021-01-29-0.json.gz'
res = requests.get(f'https://data.gharchive.org/{file}')
upload_res = s3_client.put_object(
    Bucket='itv-github',
    Key='2021-01-29-0.json.gz',
    Body=res.content
)
print(upload_res)
```

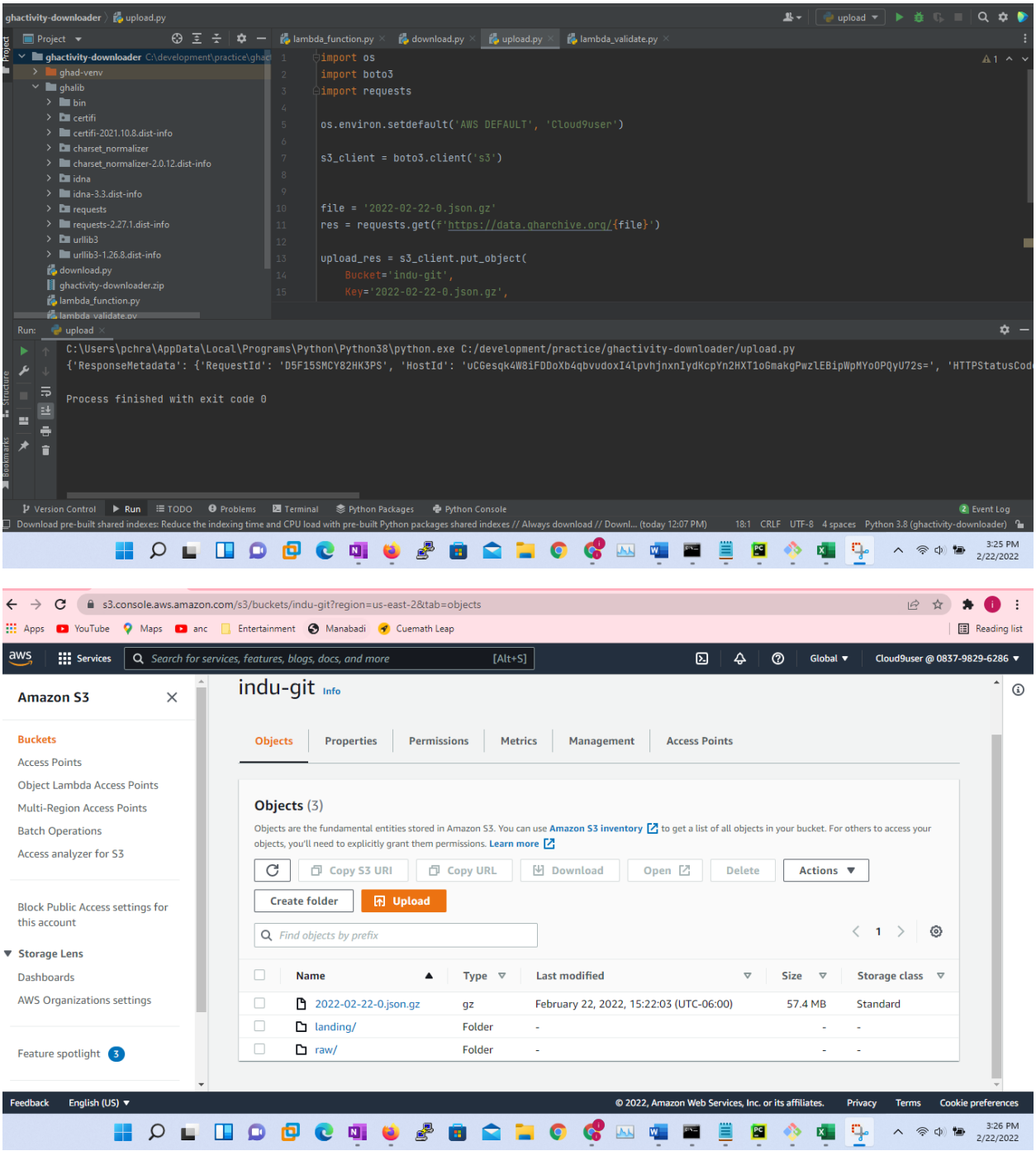


Figure 5File is uploaded into the s3 bucket from local environment

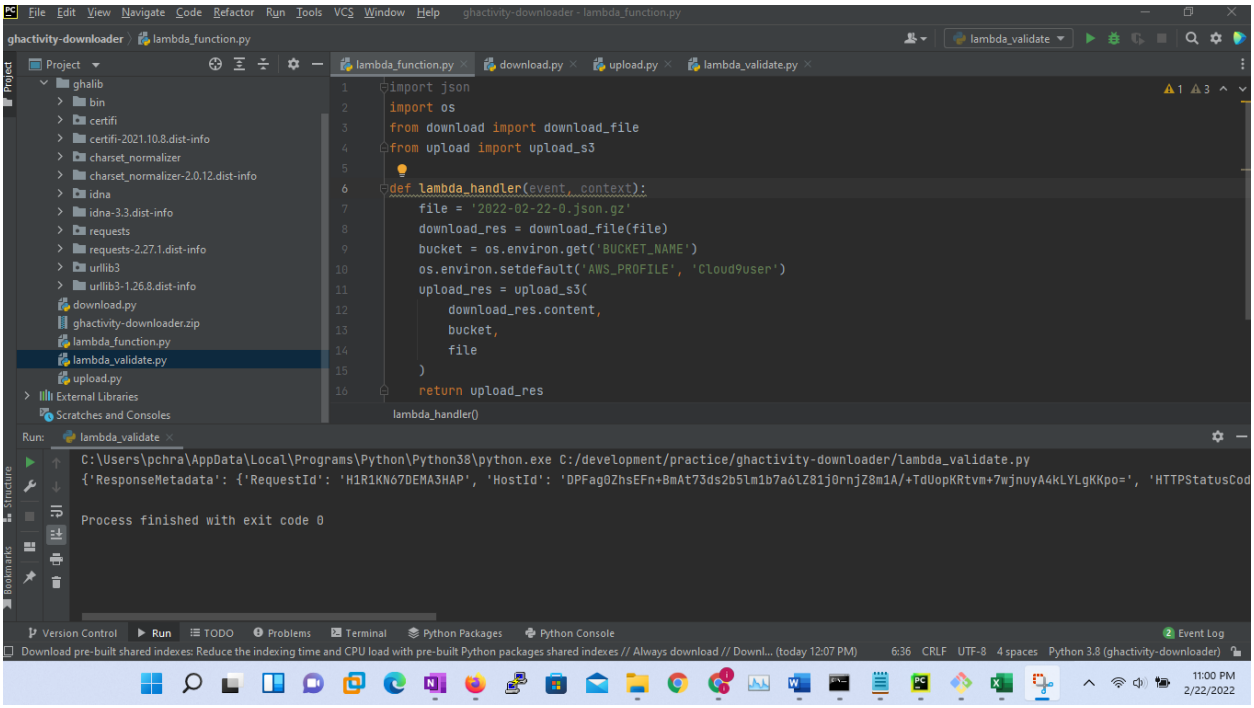
Develop upload functionality to s3:

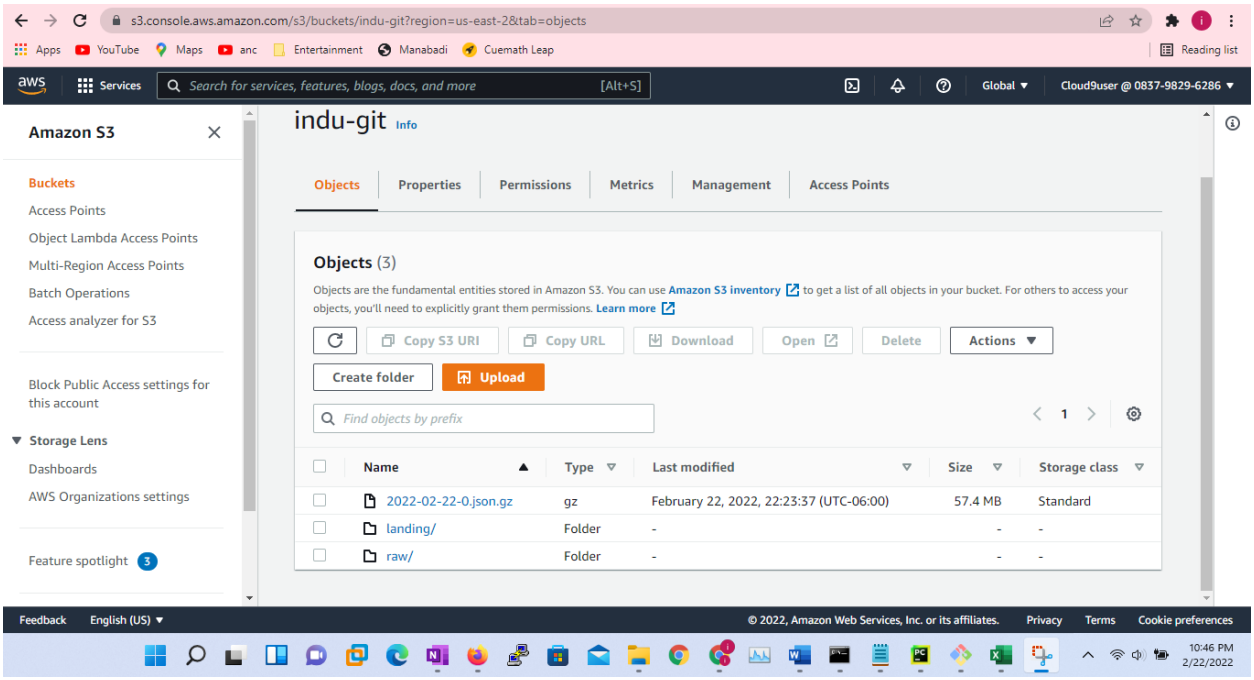
- We need to ensure that the role using which lambda function is being executed has permission on the target bucket.
- Let us create a sandbox folder under indu-git bucket.
- We need to develop a new function as part of a new program which takes the response object and uploads it as an object in s3.

```
import boto3
def get_client():
  return boto3.client('s3')
def upload_s3(body, bucket, file):
  s3_client = get_client()
  res = s3_client.put_object(
    Bucket=bucket,
    Key=file,
    Body=body
  )
  return res
```

- We also need to update the logic in the lambda handler to call the function which uploads the response as an object in s3.

```
import os
from download import download_file
from upload import upload_s3
def lambda_handler(event, context):
    file = '2021-01-29-2.json.gz'
    download_res = download_file(file)
    bucket = os.environ.get('BUCKET_NAME')
    environ = os.environ.get('ENVIRON')
    if environ == 'DEV':
        print(f'Running in {environ} environment')
        os.environ.setdefault('AWS_PROFILE', 'itvgithub')
    upload_res = upload_s3(
        download_res.content,
        bucket,
        file
    )
    return upload_res
```

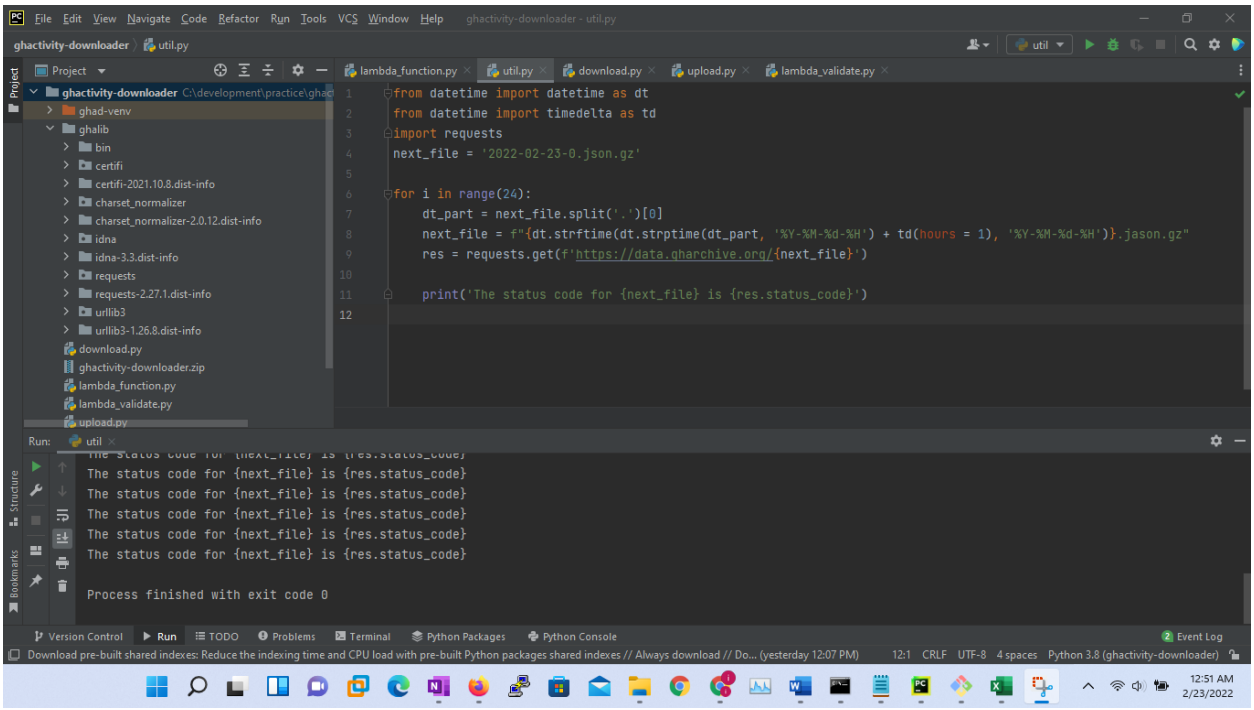




Validating files incrementally

- We can convert the date and hour part of the file to timestamp.
- Once we get the timestamp we should be able to use **timedelta** to add 1 hour in each iteration.
- Using the new date and time, we should be able to generate the next file. We can check whether the file is already available or not.

```
from datetime import datetime as dt
from datetime import timedelta as td
import requests
next_file = '2021-01-30-0.json.gz'

while True:
    res = requests.get(f'https://data.gharchive.org/{next_file}')
    if res.status_code != 200:
        break
    print(f'The status code for {next_file} is {res.status_code}')
    dt_part = next_file.split('.')[0]
    next_file = f"{dt.strftime(dt.strptime(dt_part, '%Y-%M-%d-%H') + td(hours=1), '%Y-%M-%
d-%-H')}.json.gz
```



Reading and Writing Bookmark using s3:

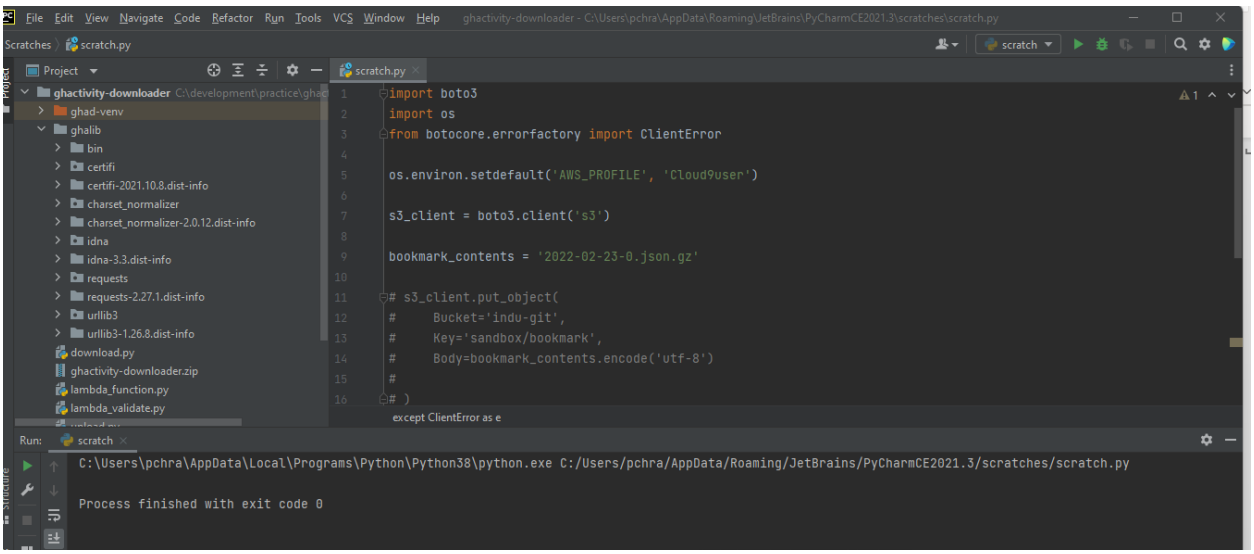
Let us go through s3 APIs to read and write bookmark details using s3. For now we will maintain the last copied file as a bookmark.

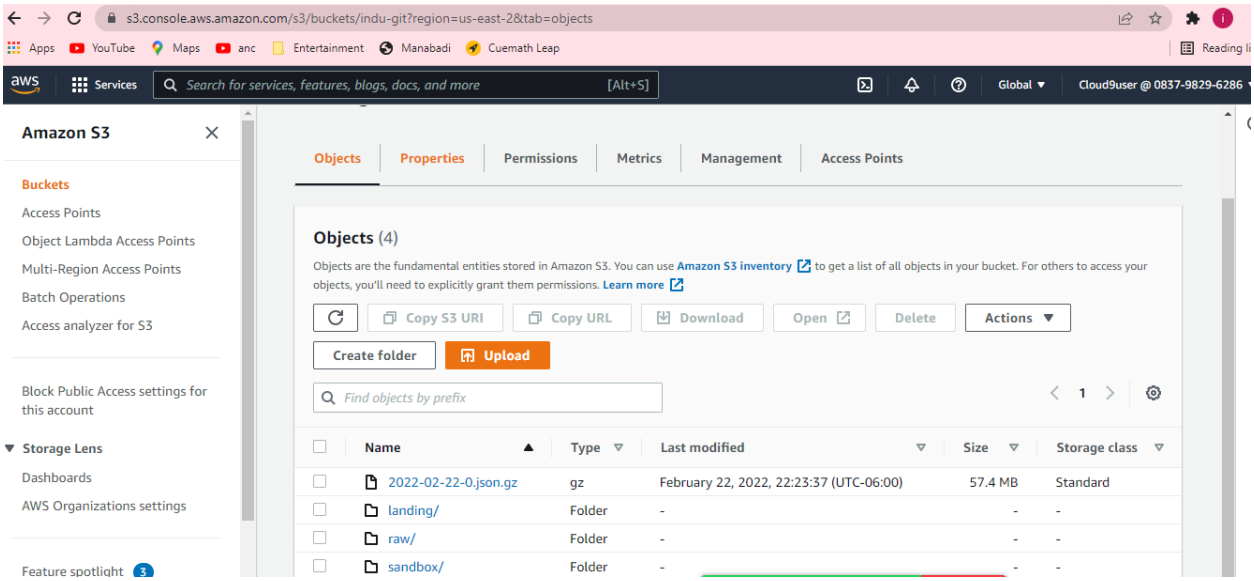
- Writing content to s3. We will use **s3://itv-github/sandbox/bookmark**.

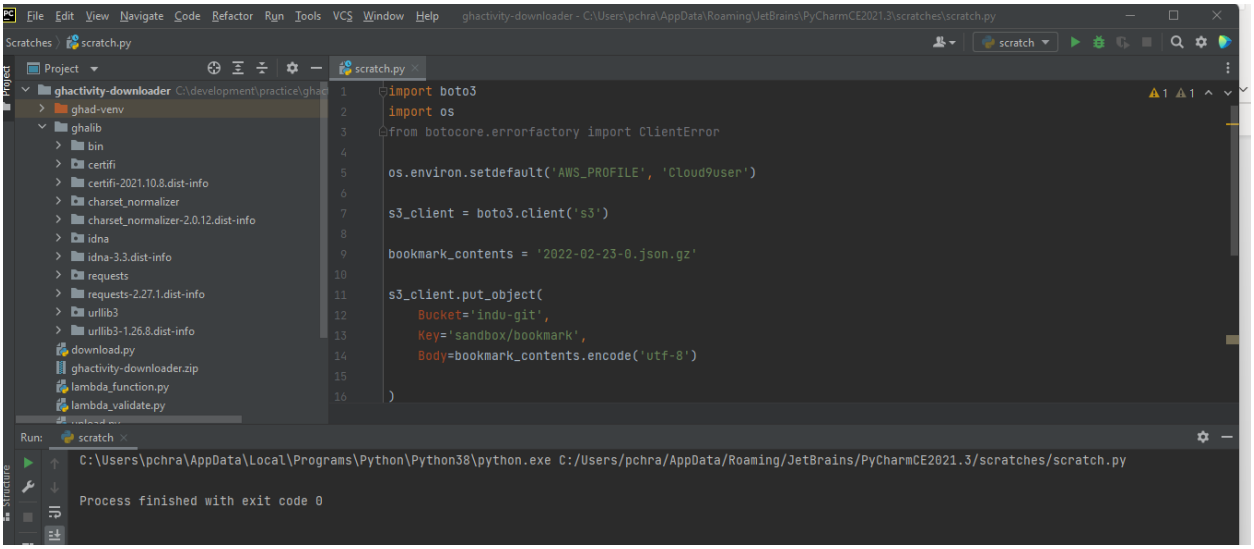
```
bookmark_contents = '2021-01-30-0.json.gz'
s3_client.put_object(
  Bucket='indu-git',
  Key='sandbox/bookmark',
Body=bookmark_contents.encode('utf-8')
)
```

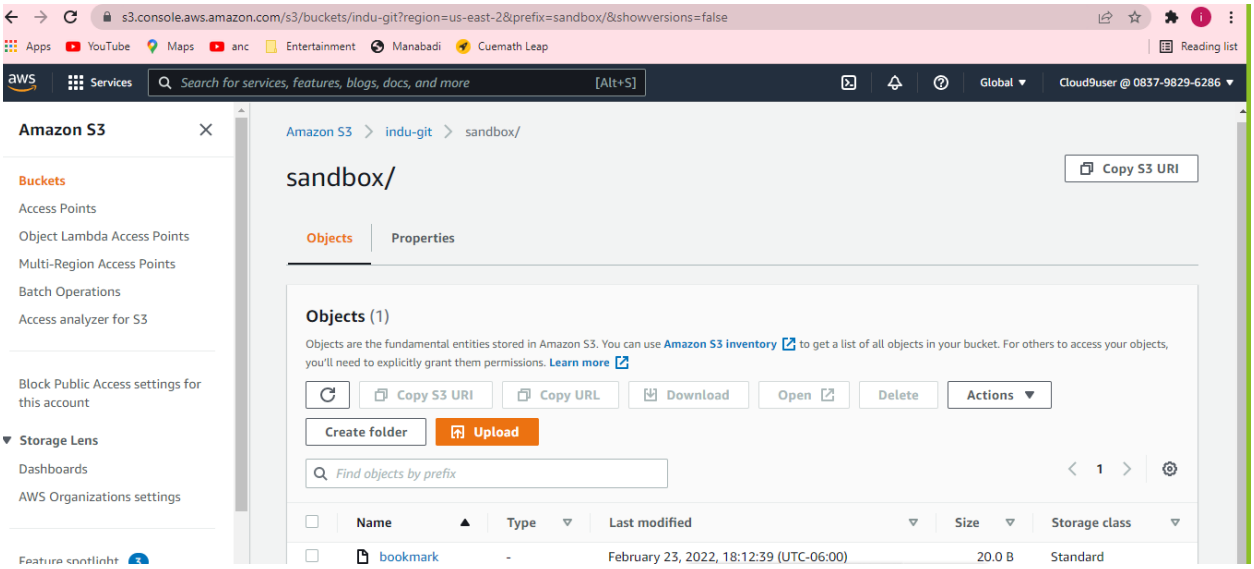
- Reading content from s3. If the bookmark is not there we need to catch the exception and use the baseline date as a bookmark.

```
try:
    bookmark_file = s3_client.get_object(
        Bucket='indu-git',
        Key='sandbox/bookmark'
    )
  prev_file = bookmark_file['Body'].read().decode('utf-8')
except ClientError as e:
    if e.response['Error']['Code'] == 'NoSuchKey':
        # Catch exception
        # prev_file = baseline_file
    else:
        raise
```






File Edit View Navigate Code Refactor Run Tools VCS Window Help
ghactivity-downloader - C:\Users\pchra\AppData\Roaming\JetBrains\PyCharmCE2021.3\scratches\scratch.py
Scratches
scratch.py
Project
ghactivity-downloader C:\development\practice\ghac
ghad-venv
ghalib
bin
certifi
certifi-2021.10.8.dist-info
charset_normalizer
charset_normalizer-2.0.12.dist-info
idna
idna-3.3.dist-info
requests
requests-2.27.1.dist-info
urllib3
urllib3-1.26.8.dist-info
download.py
ghactivity-downloader.zip
lambda_function.py
lambda_validate.py
import boto3
import os
from botocore.errorfactory import ClientError
os.environ.setdefault('AWS_PROFILE', 'Cloud9user')
s3_client = boto3.client('s3')
bookmark_contents = '2022-02-23-0.json.gz'
s3_client.put_object(
Bucket='indu-git',
Key='sandbox/bookmark',
Body=bookmark_contents.encode('utf-8')
)
Run:
scratch
C:\Users\pchra\AppData\Local\Programs\Python\Python38\python.exe C:/Users/pchra/AppData/Roaming/JetBrains/PyCharmCE2021.3/scratches/scratch.py
Process finished with exit code 0


s3.console.aws.amazon.com/s3/buckets/indu-git?region=us-east-2&prefix=sandbox/&showversions=false
Apps YouTube Maps anc Entertainment Manabadi Cuemath Leap
Reading list
aws Services
Search for services, features, blogs, docs, and more [Alt+S]
Global
Cloud9user @ 0837-9829-6286
Amazon S3
Buckets
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
Access analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
AWS Organizations settings
Amazon S3 > indu-git > sandbox/
sandbox/
Copy S3 URI
Objects
Properties
Objects (1)
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Find objects by prefix
Name
Type
Last modified
Size
Storage class
bookmark
-
February 23, 2022, 18:12:39 (UTC-06:00)
20.0 B
Standard