Module1

Important Notes:

- This module assumes you have a solid understanding of AWS and its services, python programming language and using the boto3 library to manipulate AWS resources with python code
- This module contains a single file "LambdaScraperCode.py"
- The code in this file is meant for a <u>serverless lambda function</u> in AWS
- The <u>function</u> uses two 3rd party python packages (see Packaging Guide 1)
- The <u>function</u> uses boto3 clients (see IAM Guide)
- The <u>function</u> needs access to an S3 bucket (see Bucket & Email Guide)
- The <u>function</u> needs access to an SES identity (see Bucket & Email Guide)
- The <u>function</u> needs access to the SSM parameters (see Params Guide)
- I advise you to make sure that all AWS resources that are being used to run the <u>lambda function</u> are in the same AWS region as the <u>function</u>
- Follow all the guides below in order, one by one, and hopefully you won't run into any problems

Packaging Guide 1:

1. Create a new empty folder
2. Open command prompt in that folder
3. Run the command "pip install --target . --platform linux_x86_64 --only-binary=:all: beautifulsoup4 requests"
4. Add a new file "lambda_function.py"
5. Copy and save the code from "LambdaScraperCode.py" to "lambda_function.py"
6. Select everything in the folder and add them to a new zip file

Packaging Guide 2:

1. In your AWS account create a new lambda function with python as the programming language (I recommend all default settings)
2. Configure a test event (you many use default key value pairs)
3. In "Configuration" -> "General configuration", edit "Memory", "Ephemeral storage" and "Timeout" to the highest value possible (because why not)
4. In "Code" -> "Upload from", click on ".zip file"
5. Upload the zip file you created in "Packaging Guide 1"
6. Once loaded, scroll to the top of the page and click on "+ Add trigger"
7. Select "EventsBridge" as the source
8. Select "Create a new rule"
9. Add your desired name in "Rule name"
10. Under "Rule type" select "Schedule expression"
11. Under "Schedule expression" add the rule "rate(4 minutes)", this should run your code every 4 minutes, 24/7
12. Click Add
13. You function is not ready yet. It needs many edits. Keep it open in a separate tab and follow the guides below

IAM Guide:

In "IAM" you need to create separate 3 users/users groups, each with one of the following AWS managed policies:

1. AmazonS3FullAccess
2. AmazonSESFullAccess
3. AmazonSSMFullAccess

Generate an "Access key" and "Secret access key" each time and copy and store it somewhere safe

Bucket & Email Guide:

1. In "S3" create a new bucket with all default settings
2. In "SES" create a new email address identity and then go to your inbox to verify that identity
3. Copy and store the bucket name and the email address in the same place you stored the access keys and secret access keys from the previous guide

Params Guide:

In "Systems Manager" -> "Parameter Store" you need to create three parameters with name and initial value in the table below:

Parameter name	Parameter initial value
param1	https://www.zameen.com/Homes/Karachi_DHA_Defence-213-1.html
param2	0
param3	https://www.zameen.com/Homes/Karachi_DHA_Defence-213-1.html https://www.zameen.com/Plots/Karachi_DHA_Defence-213-1.html https://www.zameen.com/Homes/Karachi_Bahria_Town_Karachi-8298-1.html https://www.zameen.com/Plots/Karachi_Bahria_Town_Karachi-8298-1.html https://www.zameen.com/Homes/Lahore_DHA_Defence-9-1.html https://www.zameen.com/Plots/Lahore_DHA_Defence-9-1.html https://www.zameen.com/Homes/Lahore_Bahria_Town-509-1.html https://www.zameen.com/Plots/Lahore_Bahria_Town-509-1.html https://www.zameen.com/Homes/Islamabad_DHA_Defence-3188-1.html https://www.zameen.com/Plots/Islamabad_DHA_Defence-3188-1.html https://www.zameen.com/Homes/Islamabad_Bahria_Town-383-1.html https://www.zameen.com/Plots/Islamabad_Bahria_Town-383-1.html

Packaging Guide 3:

Go to the section of the lambda function code that has the user specific variables as shown below:

```python
#user specific variables
S3 = client(
    "s3",
    aws_access_key_id = "your access key id",
    aws_secret_access_key = "your secret access key"
)
SES = client(
    "ses",
    aws_access_key_id = "your access key id",
    aws_secret_access_key = "your secret access key",
    region_name = "the region name"
)
SSM = client(
    "ssm",
    aws_access_key_id = "your access key id",
    aws_secret_access_key = "your secret access key",
    region_name = "the region name"
)
bucket = "your bucket"
email = "your email"
```

From the place where you stored the keys (see IAM Guide) and other things (see Bucket & Email Guide), you need to edit the variables above with your own credentials.

Once you have done this, save and deploy the edited code.

Congratulations! You have successfully installed Module1