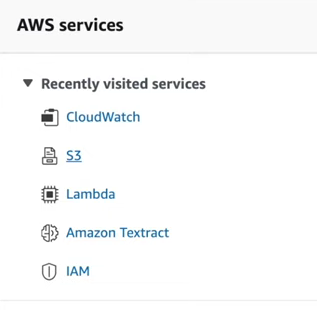
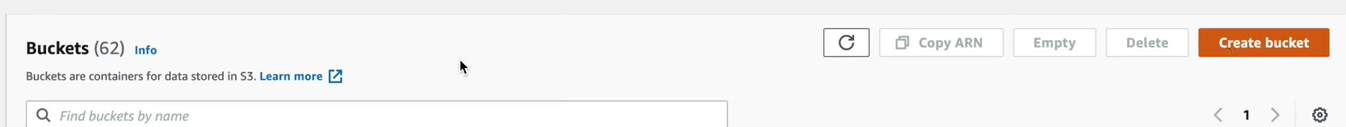
Navigate to the S3 storage console:

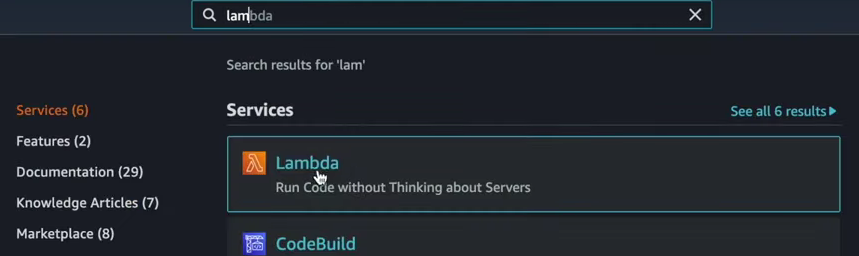


Create a new bucket:

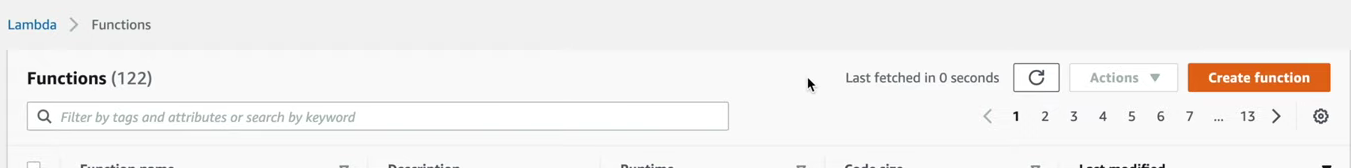


Name the bucket and then click Create bucket.

Next, navigate to the Lambda service:



Click Create Function:



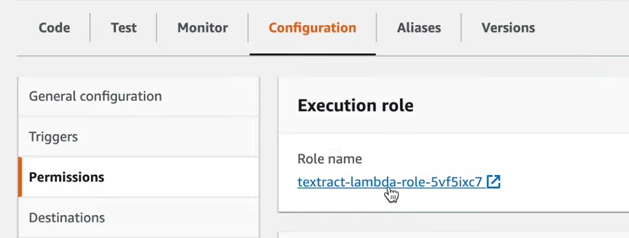
Then give the function a name and select a Python runtime (probably latest).

Then select Create a new role with basic lambda permissions.

Then click create function.

Next click on configuration under the function you just created.

Under the permissions tab, click the role name of the new lambda function. This should take you to the IAM management console for this role:

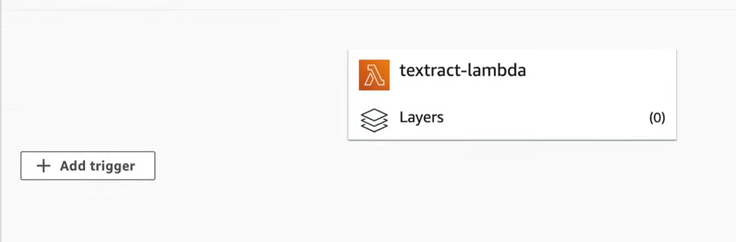


Under policies, we will attach new policies called:  




You can now close the IAM management console.

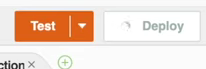
Back at the Lambda function page, we will add a new trigger:



Here we will select S3 as the trigger, find the name of the bucket and add it under bucket. Then leave the rest of the options as it is. Check the final checkbox (regarding recursive invocation) and then click add.

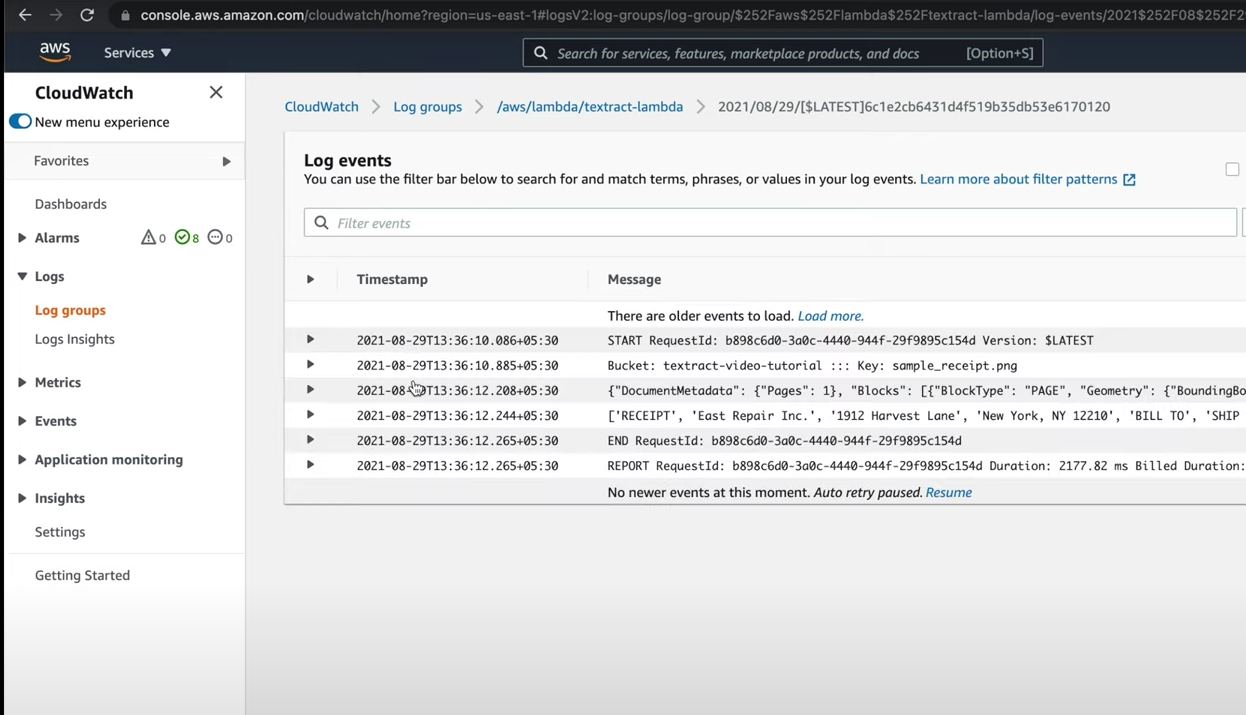
Under the code section of the lambda page for this function, Copy code from [this repository](https://github.com/srcecde/aws-tutorial-code/blob/master/lambda/lambda_textract_detect_text.py) and paste it here in place of the auto generated code:

Then Save and Deploy:



Once you go back to the Amazon S3 (the specific bucket you made) you will be able to add the file to be scanned. This will automatically trigger the lambda protocol.

With the current implementation, the raw data read from the file is not saved or sent anywhere. However, we can check the success of this feature under cloudwatch > log groups > lambda function directory. here is an example:



Additionally, this is the implementation for raw text extraction, based on how the receipt or invoice is formatted, we will be able to apply different extraction techniques for the data.

Furthermore, this data flow can be customized depending on the use case. For example, in our use-case, we will need to make some modifications to the way that files are uploaded to the S3 storage so that files can be clearly identified for specific users. This needs to be done in the case of concurrent users uploading files and making sure we know which user’s data belongs to which user.

For more information about this specific raw text extraction method implementation or details about how the output data is formatted, please refer to this video: [link](https://www.youtube.com/watch?v=lOr52VcAvzY&list=PL5KTLzN85O4LpL7cWsFHFDsKaXtrbBgPi&index=2)