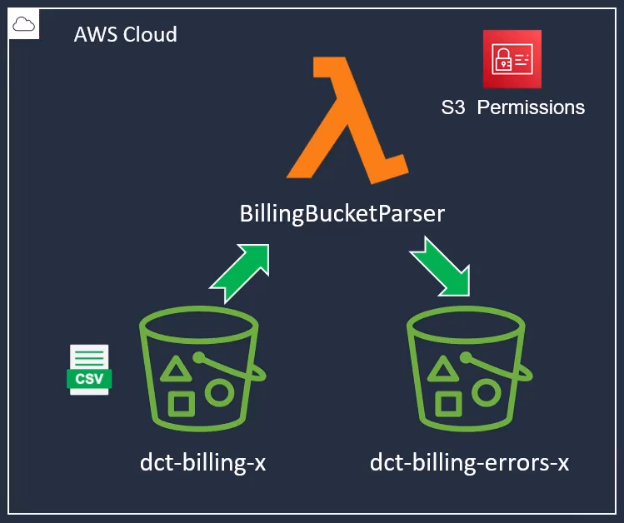
**AUTOMATING AWS WITH PYTHON | PART 5 : S3 REAL TIME VALIDATION**



Today we will setup a Lambda function which will be triggered whenever a new CSV file is uploaded to an S3 bucket. It will then validate the file and if any discrepancies are found, it will move the file to an “error” bucket.

**GitHub repository**: <https://github.com/Silas-cloudspace/python-for-aws/tree/main/3%20-%20Automating%20AWS%20with%20Lambda/2.S3_real_time_validation>

1. **Create a Lambda function:**

* Go to the AWS console, Lambda, and create a Lambda function
* Choose "Author from scratch".
* In the Function name field, enter BillingBucketParser.
* For the Runtime, select Python 3.12

1. **Create two S3 buckets**

* Create two S3 buckets, one for uploading csv files and other for error files.
* touch s3\_buckets.py
* Paste into it the code from the GitHub repository.
* Run: python s3\_buckets.py

1. **Create three files in VS code:**

* touch lambda\_function.py template.yaml event.json
* Paste into them the codes from the GitHub repository.
* Run: python lambda\_function.py

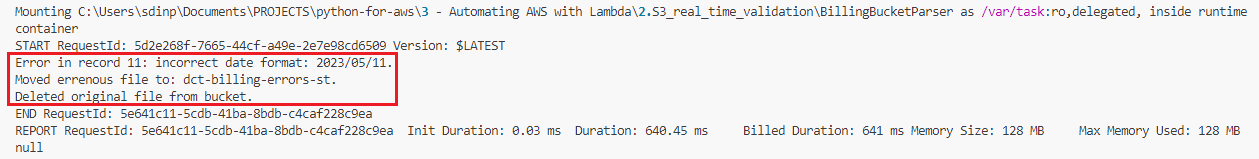
1. **File upload**

* Upload to the bucket named “dct-billing-st” the provided csv file named “billing\_data\_dairy\_may\_2023”

1. **Check the operation manually**

* On VS Code run with command prompt: sam local invoke -e event.json

\*You need to have docker installed and running



* Now you can see the results. Since an error was fond in the date format, this file was moved into the error bucket, and the original file that was in the first bucket was delete.
* Check it out in the S3 buckets on the AWS console.

1. **Create a zip file**

* Go to powershell and run: “Compress-Archive -Path lambda\_function.py -DestinationPath lambda\_function.zip”

1. **Update the Lambda function permissions**

* touch update\_policies.py
* Paste into it the code from the GitHub repository.
* Run: python update\_policies.py

1. **Increase Lambda function timeout**

Create a new file:

* touch increase\_lambda\_timeout.py
* Paste into it the code from the GitHub repository.
* Run: python increase\_lambda\_timeout.py

1. **Add a trigger to the Lambda function**

Crete a new file:

* touch lambda\_trigger.py
* Paste into it the code from the GitHub repository.
* Run: python lambda\_trigger.py

1. **Update the Lambda function**

* touch update\_lambda\_function.py
* Paste into it the code from the GitHub repository.
* Run: python update\_lambda\_function.py

1. **Test it out**

Upload the “billing\_data\_meat\_may\_2023” file into the “dct-billing.st” bucket

It should be moved into “dct-billing-errors-st” bucket automatically.