



[Exercise] Lambda Function for Data Ingestion

Data Engineering Diploma

Content developed by: WeCloudData Academy

In this exercise, you need to create a Lambda function to:

1. download 'inventory.csv' from a public AWS S3 URL.
2. upload the 'inventory.csv' file to the Snowflake table.
3. create a schedule for every night 2am (EST) time.

1. Download file from S3

The URL of the S3 bucket is '<https://de-materials-tpcds.s3.ca-central-1.amazonaws.com/inventory.csv>', it is fixed, but the inventory file you are going to download is updated every day.

You need to use the library of `requests`, and also, in the later step, you also need an external library `snowflake-connector-python`. So you need to add a new layer for the Lambda Function. Below is a quick way to add a layer.

In []: *# you can make this layer on your CloudShell or EC2 instances. But CloudShell may requires some new installation of packages.*

```
mkdir -p lambda_layers/python/lib/python3.7/site-packages
virtualenv venv
python3 -m venv venv
source venv/bin/activate
# 1) install the dependencies in the desired folder
pip3 install -r requirements.txt -t lambda_layers/python/lib/python3.7/site-packages/
# 2) Zip the Lambda_Layers folder
cd lambda_layers
zip -r snowflake_lambda_layer.zip *
# 3) publish Layer
aws lambda publish-layer-version \
  --layer-name fl-snowflake-lambda-layer \
  --compatible-runtimes python3.7 \
  --zip-file fileb://snowflake_lambda_layer.zip
```

You need to read the URL, and download the file to `tmp/inventory.csv` ('tmp' folder is the only folder you can use save files temporarily on Lambda)

2. Upload the file to Snowflake

This step is harder than the first step, because it require multiple operations. Before the Lambda function, make sure you have made schema and table on the snowflake in advance.

In the Lambda function, you need:

1. Tell Snowflake which Schema you are going to use;
2. Create File Format for the file;
3. Create a stage;
4. Upload file from the lambda 'tmp' folder to the Snowflake stage;
5. Copy the file from the stage to the table in the Snowflake.

3. Schedule

You need to set a EventBridge in the LambdaFunction to make it run on every night 2 am (EST).

4. Help

You can find help from [this code](#). And also, consider use ChatGPT to help you finish the code.