Data Engineer Bootcamp (Full-Time)

HM
HIBAHMOHAMMED O SINDI
haboba1417@hotmail.com

WeCloudData

Data Engineering Diploma

Content developed by: WeCloudData Academy

Mandatory: Yes
Group: Yes
In Class: Yes
Submission:No
Grading: No
Demo:No
Duration: 4 hrs

In this lab time, let's create an AWS Lambda function. In this note, 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 2 am (Riyadh 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.

```
# you can make this layer in Cloudshell

# Open Cloudshell

# create layer

mkdir -p lambda_layers/python/lib/python3.9/site-packages
python3 -m venv venv
source venv/bin/activate
# 1) install the dependencies in the desired folder
# You will realize that we are not using requirements.txt file here and pip3 install command is also using multiple flags. This is because lambda backend c
# Please review this link (https://stackoverflow.com/questions/75472308/aws-lambda-returns-lib64-libc-so-6-version-glibc-2-28-not-found) for more informati
pip3 install --platform manylinux2010_x86_64 --implementation cp --python 3.9 --only-binary=:all: --upgrade --target lambda_layers/python/lib/python3.9/sit
# 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 python3.9 \
    --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 to save files temporarily on Lambda).

2 Upload the file to Snowflake

This step is harder than the first step because it requires 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 the file from the Lambda 'tmp' folder to the Snowflake stage;
5. Copy the file from the stage to the table in Snowflake.

3 EventBridge setting

You need to set an EventBridge in the LambdaFunction to make it run every night at 2 am.

4. Testing

Testing the Lambda to see if it works.

5 Help

You can find help from this code . And also, consider use ChatGPT to help you finish the code.

Course Content
Enter code

All
Lecture
Recordings
Practices
1
Program Information
Chapter overview
Program Administration
Grading and Attendance
About the projects in the bootcamp
How to use the Learning Portal
Daily Schedule
2
Surveys
Chapter overview
Surveys
Week 0 Survey - Student Background
Week 3 Survey - Client Project
Week 4 Survey
Project Group Survey
3
Week 00 (Virtual)- Program Preparation
Chapter overview
Week Plan
Week Plan
Software Installation
[Software Installation]: VsCode
[Software Installation]: Jupyter notebook
[Software Installation]: Python
[Software Installation]: MySQL
[Software Installation]: Unbuntu on Mac
[Software Installation] Ubuntu on Windows
[Online Platform Use]: Colab
Pre-bootcamp
Pre-bootcamp Material

Presentations

[Lecture Video] Sunday: Orientation Session

[Lecture Video] Tuesday: Introduction to Data Engineering

[Lecture Video] Wednesday: Curriculum

[Lecture Video] Thursday: Curriculum

[Lecture Slide] (Wed) Curriculum Introduction

4

Week 01 - SQL

Chapter overview

Sunday - Basic SQL

[Lecture Materials]SQL basics

[Lab] SQL Basics Exercise

[Lab] Exercise: SQL - Airbnb (Optional)

[Lecture video] SQL Basics

Monday - SQL Join and sub-select

[Lecture Materials] SQL join and sub-select

[Lab] Exercise: Join and Sub-select

[Lecture video] SQL Day 2

Tuesday - SQL Window Function

[Lecture Materials] SQL Window Function

[Lab] Exercise: Window Function

[Lecture video] SQL Day 3

[Lab Video] SQL Lab Solution

Wednesday - SQL DDL and CTE

[Lecure Materials] DDL and CTE

[Lab] SQL ddl

[Lab] SQL CTE

[Lecture Video] SQL Wednesday

Thursday - SQL Review

[Weekly Quiz] SQL - Week 1

[Lecture Video] SQL Thursday Review

[Lecture Slides] SQL Review

5

Week 02 - Python

Chapter overview

Sunday - Python data type and structure

[Lecture Materials] Python Data Structure and Data Types

[Lab] Exercise: Python Data type and structure

[Lab] Exercise: OpenAI ChatBot (Optional)

[Lecture Video] Python Sunday

Thursday - Holiday

Monday - Python Control Flow and Function

[Lecture Material] Python Control Flow and Function
[Lab] Exercise: Python Function
[Lab] Exercise: Python Control Flow
Python Quiz (Multiple-Choice)
[Lecture Video] Python Monday
Tuesday - Pandas 1
[Lecture Material] Pandas 1
[Lab] Pandas Intro
[Lecture Video] Python Tuesday
Wednesday - Pandas 2
[Lecture Material] Pandas 2
[Lab Demo] PandaSQL
[Lab] Exercise: Advanced Pandas
[Lecture Video] Python Wednesday
6
Week 03 - Client Project
Chapter overview
Sunday - Real Client Project Intro
[Lecture Material] Web Scraping
[Real Client Project] Project Requirements
[Note] Project Group Assignment
[Lecture Video] Webscraping Sunday
Monday - Real Client Project Day
Tuesday - Real Client Project
[Real Client Project] Code & Data Submission
Wednesday - RCP
Thursday - RCP
[Lecture Video] Webscraping Thursday
7
Week 04 - Linux and AWS
Chapter overview
Sunday - Linux
[Lecture Material] Linux
[Lab] Exercise: Bash Commands
[Lab] Mini Project: Riyadh Climate Data - Cron Job
[Lecture Video] - Linux Sunday
Monday - AWS Intro
[Lecture Material] AWS Intro
[Lab] AWS Account Setup
[Lab] Workshop AWS EC2
[Lab] Workshop S3
[Lecture Video] AWS Monday
Tuesday - Lambda

[Lab] Workshop: Lambda
[Lecture Material] Lambda
[Lab] Mini Project: Lambda
[Lecture Video] Lambda Tuesday
Wednesday - Practice Day
[Lecture Material] Plan For Today
[Lab video] 2024-03-06
Thursday - Practice Day
[Lecture Material] Plan For Today
[Quiz] Linux and AWS Quiz
[Lab Video] EC2, S3, Lambda workshops demo
8
Week 05 - Docker and Client Project phase 2
Chapter overview
Sunday - Docker I
[Lecture Material] Docker
[Lab] Software Installation: Docker
[Lab] Account Creation Create your Dockerhub account
[Lab] Workshop Demonstrating Hello World Example
[Lab] Workshop: Work with Docker Image
[Lab] Exercise: Basic Docker Commands
[Lecture Video] Docker Sunday
[Lab] Exercise: Basic Docker Commands Updated
Monday - Docker II
[Lab] Workshop: Install Zepplin with Docker
[Lab] Workshop: Docker Compose --Flask
[Quiz] Docker Commands Quiz
[Lecture Video] Docker II - Monday
[Lab] Workshop: Install Zepplin with Docker Updated
Tuesday - Real Client Project Phase 2
[Lecture Video] Learning Roadmap & RCP Feedback
Wednesday - Real Client Project Phase 2
Thursday - Real Client Project Phase 2
RCP project Submission (Competition)
9
Week 06 - Data Warehouse
Chapter overview
Sunday - Snowflake Data Warehouse
[Lecture Material] Snowflake
[Lab- W601]: Software Installation: DBeaver
[Lab-W602]: Account Creation: Snowflake
[Lab-W603]: Software: Connect Snowflake with DBeaver

[Lab-W604]: Exercise: Snowflake
[Lecture Video] Snowflake - Sunday
Shaohua Weekly Review [RCP] - Sunday
[Lab Video] Snowflake Demo - Monday
Monday - Data Warehouse Intro
[Lecture] Data Warehouse Intro
[Quiz-W611] Data Warehouse Concept (Grading!!)
[Lab-W612] Exercise: Use SnowSQL (Optional)
[Lecture Video] Data Warehouse - Monday
Tuesday - SQL in ETL
[Lecture Materials] SQL in ETL
[Lab] TA Exercises Review
[Lecture Video] SQL in ETL - Tuesday
Wednesday - Data Modeling and ETL
[Lecture Material] Data Modeling and ETL
[Lab-W631] Exercise: Data Modelling and ETL (Group)
[Lecture Video] Data Modelling and ETL - Wednesday
Thursday - Data Loading
[Lecture Material] Data Loading
[Lab-W641] Exercise: ETL and Data Loads (Group)
[Lecture Video] Data Loading
10
Week 07 {Project Week} - Capstone Project-1
Chapter overview
Sunday - Data Warehouse Review
[Lecture Material] Agenda For Today
[Lecture Video] Data Warehouse Lab Review - Sunday
Monday - {Capstone Project} Project Intro
[Project Material] Project Guideline
[Project Material] Project Data Overview
[Project Material] Business Requirements Overview
[Project Material] Project Infrastructure Overview
[Lab-W711] Project Task1: Setup Snowflake, EC2 and Docker
[Project Material] Project Introduction (Full-version)
[Lecture Video] Capstone Project Intro - Monday
Tuesday - {Capstone Project} Lambda Setup in Project
[Lab-W721] Project Task2: AWS Lambda Setup
[Lecture Video] Capstone Project Lambda Setup - Tuesday
Wednesday - {Capstone Project} Airbyte Setup in Project
[Lab-W731] Project Task3: Airbyte Installation and Configuration
Thursday - {Capstone Project} Self-work On Project Part 1