WeCloudData

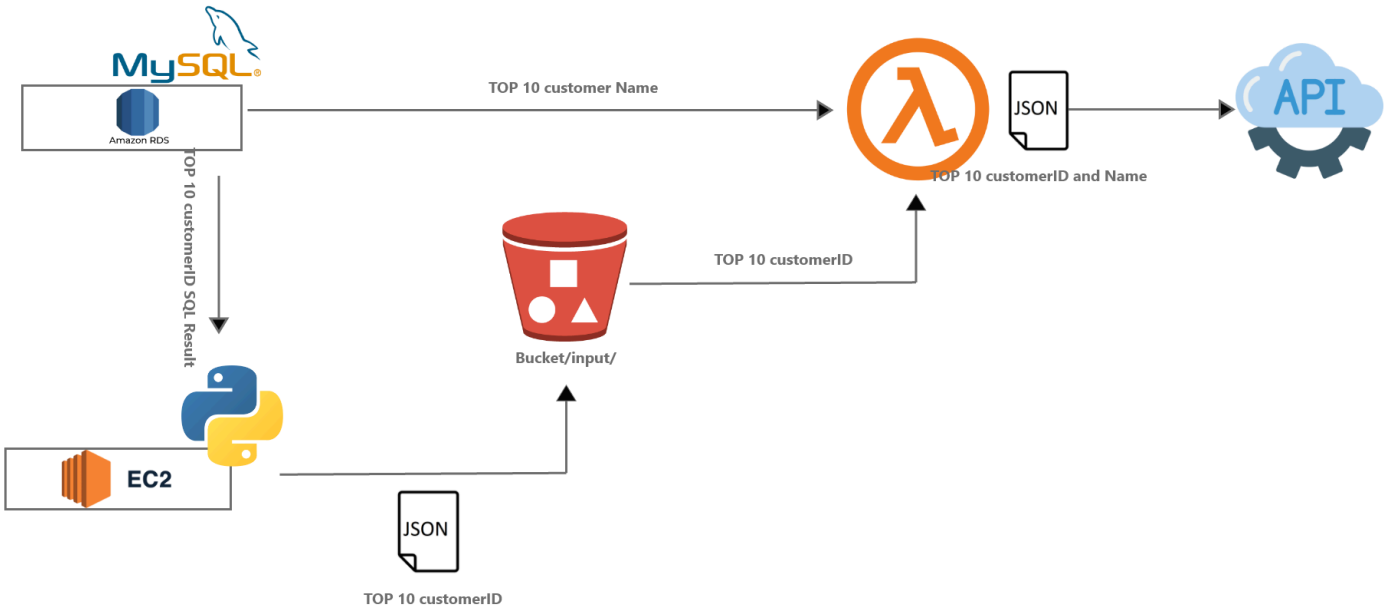
Lab 2--Python Cloud Project (Lambda)

Data Engineering Diploma

Content developed by: WeCloudData Academy

1. Project Description

In this project, you need to read data from a relational database (MySQL), save the result to S3, and finally post your result to an API endpoint. Here is the general architecture:



2. Detail Steps

1. Start a MySQL database on AWS RDS and connect RDS with MySQL Workbench. If you don't know how to set this, please watch the (video.)
2. Download the sql script from the **DOWNLOADS**. And run the script on MySQL to load data. You will have a database called **superstore**.
3. Query the database, to get the top 10 customer ids who have the most purchase. Get their **customer id** and **sum of the customer sales** (you can use different column name) from the database.

4. After getting the result, save the result to S3. In order to do so:

    - 1) You need to create an EC2 instance and build a python project.
    - 2) Use the python script to bring the result from MySQL to EC2. The result includes the **customer id** and the **sum of the customer sales**of the top 10 customers.
    - 3) Save the result as a **.json** file locally in EC2.
    - 4) Upload the file from EC2 to S3. The file should be in an 'input' folder in the S3 bucket.

5. When the file lands on the S3 bucket, a lambda is triggered. The lambda function will:

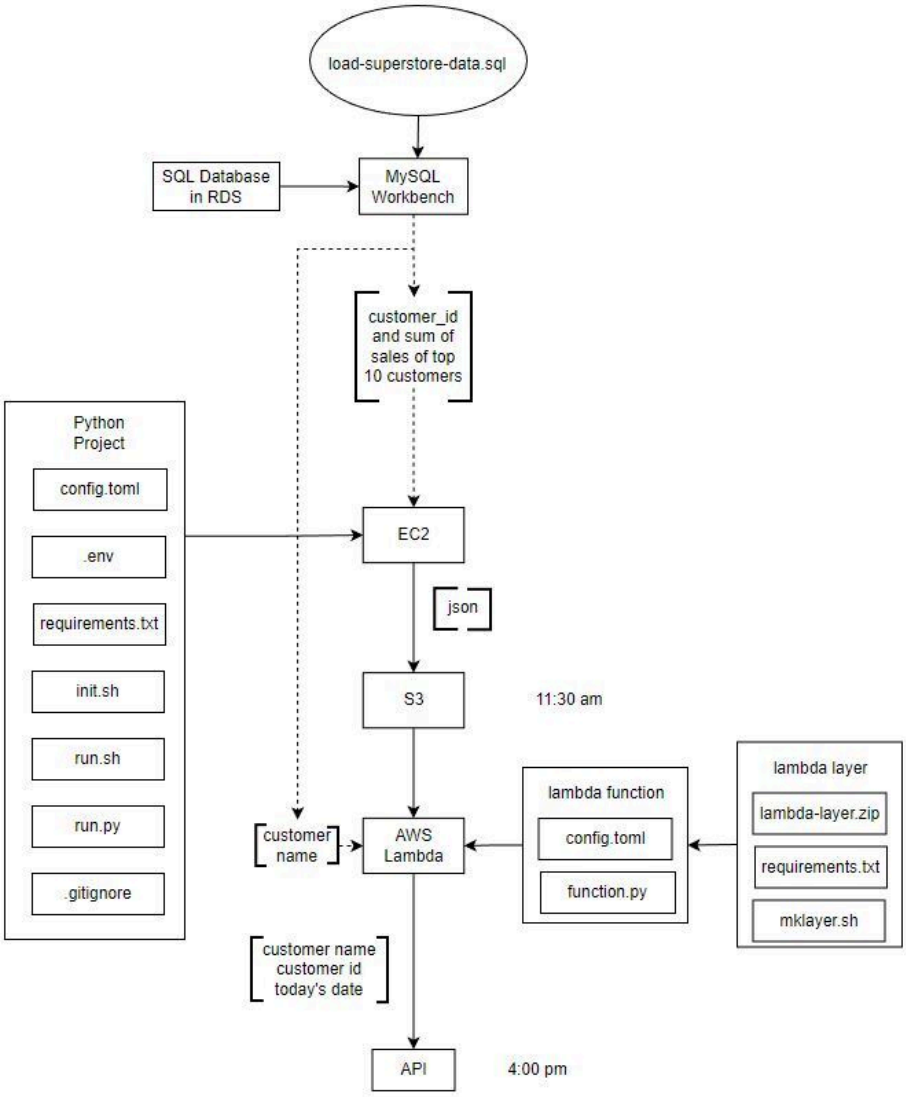
    - 1) get the **customer id** list from the S3 **..json** file;
    - 2) query the customer names from the database based on the **customer id** list.
    - 3) send a JSON data, including **customer id** , **customer name** and **today's date** to an API endpoint. The API endpoint is: https://virtserver.swaggerhub.com/wcd_de_lab/top10/1.0.0/add
    - 4) Here are some tips for the lambda function:
        - When you create a lambda function, you may need a lambda layer to install 3rd party libraries. This is an article about how to add a layer.(article)
        - Put the **customer id**, **customer name** and **today's date** (format: '1990-01-01') in a JSON structure variable, such as data. (But the variable is a string).
        - Use POST to send data. Here is what the final API data looks like:


```
[
  {
    "id":91811409,
    "name":"Ritsa Hightower",
    "date":"2022-06-28"
  },
  {
    "id":64444841,
    "name":"Becky Pak",
    "date":"2022-06-28"
  },
  {
    "id":38079848,
    "name":"Raymond Book",
    "date":"2022-06-28"
  },
  {
    "id":104752832,
    "name":"Jasper Cacioppo",
    "date":"2022-06-28"
  },
  {
    "id":23905406,
    "name":"Dennis Kane",
    "date":"2022-06-28"
  },
```


        - When the **POST** Succeed, the return code should be **201**. Use the **status_code** method in requests to get the returned code.
    - 5) Use python project structure.
    - 6) Use git repository to save your code.

4. Diagram

The project diagram is below:



5. HELP

In case you are facing a big challenge to finish such project, you can refer to (this link)

Chapter
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
Wednesday - Practice Day
[Lecture Material] Plan For Today
Thursday - Practice Day
[Lecture Material] Plan For Today

[Lab] Mini Project: Lambda