

Processing for DoorDash Delivery Data

Overview:

This assignment involves creating an automated AWS-based solution for processing daily delivery data from DoorDash. JSON files containing delivery records will be uploaded to an Amazon S3 bucket. An AWS Lambda function, triggered by the file upload, will filter the records based on delivery status and save the filtered data to another S3 bucket. Notifications regarding the processing outcome will be sent via Amazon SNS.

Requirements:

- AWS Account
- Amazon S3 buckets: doordash-landing-zn and doordash-target-zn
- AWS Lambda
- Amazon SNS
- AWS IAM (for permissions)
- AWS CodeBuild (for CI/CD)
- GitHub (for version control)
- Python, pandas library
- Email subscription for SNS notifications

Steps:

- Sample JSON File for Daily Data:
 - A sample JSON file named **2024-03-09-raw_input.json** with 10 delivery records, including different statuses like cancelled, delivered, and order placed.

```
{
  "id": 1, "status": "delivered", "amount": 20.5, "date": "2024-03-09"
},
{
  "id": 2, "status": "cancelled", "amount": 15.0, "date": "2024-03-09"
},
{
  "id": 3, "status": "order placed", "amount": 22.5, "date": "2024-03-09"
},
{
  "id": 4, "status": "delivered", "amount": 19.5, "date": "2024-03-09"
},
{
  "id": 5, "status": "cancelled", "amount": 18.0, "date": "2024-03-09"
},
{
  "id": 6, "status": "delivered", "amount": 23.5, "date": "2024-03-09"
},
{
  "id": 7, "status": "order placed", "amount": 20.0, "date": "2024-03-09"
},
{
  "id": 8, "status": "delivered", "amount": 25.5, "date": "2024-03-09"
},
{
  "id": 9, "status": "delivered", "amount": 21.5, "date": "2024-03-09"
},
{
  "id": 10, "status": "cancelled", "amount": 17.5, "date": "2024-03-09"
}
```

- Daily JSON will arrive in S3 bucket **doordash-landing-zn** with format like yyyy-mm-dd-raw_input.json
- As soon as file lands in S3, data processing should start
- **Set Up S3 Buckets:**
 - Create two S3 buckets: doordash-landing-zn for incoming raw files and doordash-target-zn for processed files.
- **Set Up Amazon SNS Topic:**
 - Create an SNS topic for sending processing notifications.

- Subscribe an email to the topic for receiving notifications.
- **Create IAM Role for Lambda:**
 - Create an IAM role with permissions to read from **doordash-landing-zn**, write to **doordash-target-zn**, and publish messages to the SNS topic.
- **Create and Configure AWS Lambda Function:**
 - Create a Lambda function using Python runtime.
 - Add the pandas library to the function's deployment package or use a Lambda Layer for pandas.
 - Use the S3 trigger to invoke the function upon file uploads to doordash-landing-zn.
 - The Lambda function should:
 - Read the JSON file into a pandas DataFrame.
 - Filter records where status is "delivered".
 - Write the filtered DataFrame to a new JSON file in doordash-target-zn using the specified format.
 - Publish a success or failure message to the SNS topic.
- **AWS CodeBuild for CI/CD:**
 - Host your Lambda function code on GitHub.
 - Set up an AWS CodeBuild project linked to your GitHub repository.
 - Configure the buildspec.yml to automate deployment of your Lambda function code updates.
- **Testing and Verification:**
 - Upload the sample JSON file to doordash-landing-zn and verify that the Lambda function triggers correctly.
 - Check doordash-target-zn for the processed file and confirm its contents.
 - Ensure an email notification is received upon processing completion.