Assignment 4

CSCI 5410 (Serverless Data Processing)
Date Given: Nov 16, 2022
Due Date: Nov 29, 2022 at 11:59 pm
Since this is the last assignment, late Submissions are not accepted and will result in scoring "0" in the assignment.

To avoid any additional charges for resource consumption - Delete any AWS service and GCP resources, storage, database after fulfilling the assignment submission requirements

---

Objective:
This assignment will help you learn some key services of AWS platform, and GCP platform. In this assignment, you are required to work on AWS Lambda/SQS/SNS, Google BigQueryML

Plagiarism Policy:

- This assignment is an individual task. Collaboration of any type amounts to a violation of the academic integrity policy and will be reported to the AIO.
- Content cannot be copied verbatim from any source(s). Please understand the concept and write in your own words. In addition, cite the actual source. Failing to do so will be considered as plagiarism and/or cheating.
- The Dalhousie Academic Integrity policy applies to all material submitted as part of this course. Please understand the policy, which is available at: https://www.dal.ca/dept/university_secretariat/academic-integrity.html

Assignment Rubric - based on the discussion board rubric (McKinney, 2018)

	Excellent (25%)	Proficient (15%)	Marginal (5%)	Unacceptable (0%)	Problem # where applied
Completeness including Citation	All required tasks are completed	Submission highlights tasks completion. However, missed some tasks in between, which created a disconnection	Some tasks are completed, which are disjoint in nature.	Incorrect and irrelevant	**Part A** **Part B**
Correctness	All parts of the given tasks are correct	Most of the given tasks are correct However, some portions need minor modifications.	Most of the given tasks are incorrect. The submission requires major modifications.	Incorrect and unacceptable	**Part A** **Part B**
Novelty	The submission contains novel contribution in key segments, which is a clear indication of application knowledge.	The submission lacks novel contributions. There are some evidence of novelty, however, it is not significant	The submission does not contain novel contributions. However, there is an evidence of some effort.	There is no novelty	**Part A** **Part B**
Clarity	The written or graphical materials, and developed applications	The written or graphical materials, and developed applications do not show clear picture of	The written or graphical materials, and developed applications fail to prove the clarity.	Failed to prove the clarity. Need proper background knowledge to perform the tasks.	**Part A** **Part B**

	provide a clear picture of the concept and highlights the clarity.	the concept. There is room for improvement	Background knowledge is needed.		

Citation:
McKinney, B. (2018). The impact of program-wide discussion board grading rubrics on students' and faculty satisfaction. Online Learning, 22(2), 289-299.

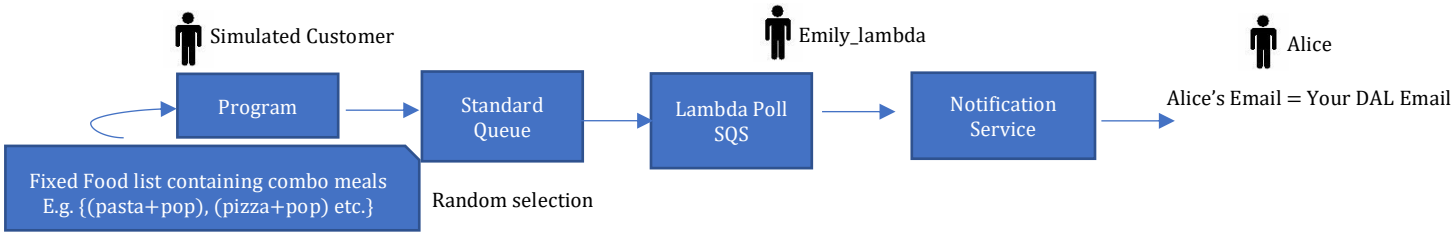
Tasks:

This assignment has 2 parts. Part A is related to coding, development, and testing. Part B is an experiment related to GCP BigQueryML

Part A. Use AWS Lambda-SQS-SNS:

take screenshots at every step and submit as part of the PDF:

a. Alice and Emily work at HfxDine, which is a cloud kitchen that provides delivery service. They receive orders online. Alice delivers food parcels, and Emily prepares it.
b. The customers' orders are added to a Queue (standard SQS) –
   - Assume a program is sending random food order messages to HfxDine
   - This message simulates how a customer places an order to an online store. (You can ignore, price/tax etc)
   - The program should randomly pick one combo meal from a list, create a message body, and send to HfxDine
c. Emily periodically (every 500 ms or any delay of your choice) checks, if there is any order in the Queue.
d. If message is available, it is assumed Emily has packed the combo meal, and then a notification service (SNS) is triggered which sends the details to Alice's email (Your email in this case)



Part A - Submission requirement:

Submit screenshots of every steps. Please do not exclude any steps. Include all screenshots as part of a PDF file. In addition, provide the program/scripts as part of the PDF file. Submit code as part of PDF and also upload the code in gitlab.

Part B. Use GCP BigQueryML:

take screenshots at every step and submit as part of the PDF:

a. Read and understand GCP BigQueryML, which you will need to create KMeans cluster of the given dataset.
   https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create-kmeans
b. Use the given dataset "SDey_FTP_input.csv" as your training dataset and test dataset to form and test the model. The dataset contains two columns, vector1 and vector2. Use Euclidean distance as the distance type to measure distance between the given vectors

c. Create a random sampling with 75% of the record that is available in the given file as training set and remaining 25% as the test set.
d. You can create the two datasets (training and test) before performing the BigQueryML operation.
e. If you need clarification about this specific problem, do not hesitate to ask me (send me email or message on Teams)

Part B- Submission requirement:

Submit screenshots of every steps. Please do not exclude any steps. Include all screenshots as part of a PDF file. In addition, provide any program/query as part of the PDF file. Submit any code you write for this problem as part of PDF and upload the code in gitlab.