Assignment 3

CSCI 5410 (Serverless Data Processing)
Date Given: Jun 13, 2022
Due Date: Jun 24, 2022 at 11:59 pm
Late Submissions are not accepted. A deduction of 10% per day will be applied for late submission.

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

---

Objective:
This assignment will help you learn a key concept related to Cloud computing through a literature study. In addition, by implementing a simple Serverless application, you will get practical experience of event-driven application building.

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 provide a clear picture of the concept and highlights the clarity.	The written or graphical materials, and developed applications do not show clear picture of the concept. There is room for improvement	The written or graphical materials, and developed applications fail to prove the clarity. Background knowledge is needed.	Failed to prove the clarity. Need proper background knowledge to perform the tasks.	**Part A** **Part B**

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, and development. Part B is related to background literature study.

Part A. Build an event-driven serverless application using AWS Lambda.

In this assignment, you need to use S3 bucket, DynamoDb, and Lambda Functions.
**[B00xxxxxx = your B00 number] used in bucket naming

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

a. Create your 1st S3 bucket **SourceB00xxxxxx** and 2nd S3 bucket **TagsB00xxxxxx** using **AWS SDK** (any programming language)
b. Upload the files given in the ***Tech*** folder **one at a time** with a delay of 200 milliseconds on the 1st bucket. You need to write a script using SDK to upload the files one at a time to the S3 bucket.
c. If a file is available on the 1st bucket, then it triggers **extractFeatures** Lambda function, which is the 1st lambda function.
d. This lambda function extracts the Named entities from the file and creates a JSON array of named entities* for that file.
e. E.g. 001.txt contains Asia, Soviet, Serbia etc., then the JSON array created by the function should be "001ne": {"Asia":1, "Soviet":1…..etc.}.
f. This file will be saved as 001ne.txt in a new bucket - **TagsB00xxxxxx**.
g. Once the file is available on this 2nd bucket, then **accessDB** Lambda function will automatically be triggered.
h. **accessDB** is your 2nd Lambda function. This Lambda function reads each named entity JSON file and updates the DynamoDb database table (three entries/array - NameEntity, Frequency, TimeStamp of Entry).

E.g. 001ne.txt contains "001ne": {"Asia":1, "Soviet":1…..etc.}. Then **accessDB** Lambda function will update DynamoDb database table where "Asia" will be a value for field "NamedEntity", and "1" will be the value for field "Frequency".

i. Test your Serverless application with test cases and provide screenshots.

Part A - Submission requirement:

For (a to h), 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. Include citation, such as documentation you followed, materials you read etc. Furthermore, upload the code on gitlab.

Additional Information:

Dataset Citation: **D. Greene and P. Cunningham. "Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering", Proc. ICML 2006.**

***Named Entity:** "In information extraction, a named entity is a real-world object, such as persons, locations, organizations, products, etc., that can be denoted with a proper name. It can be abstract or have a physical existence. Examples of named entities include Barack Obama, New York City, Volkswagen Golf, or anything else that can be named." – Wikipedia
Hint: Any word that starts with an uppercase or all letters of the word is in uppercase, then it can be considered as a named entity. E.g. Saurabh, IBM, DAL, Halifax, Serverless etc.

Part B.

I. Indu, P.M. Rubesh Anand, Vidhyacharan Bhaskar, "Identity and access management in cloud environment: Mechanisms and challenges", *Engineering Science and Technology, an International Journal*, Volume 21, Issue 4, 2018, Pages 574-588, ISSN 2215-0986, https://doi.org/10.1016/j.jestch.2018.05.010

Indu et al. published the above paper on identity and access management in cloud computing, which highlights some key concepts on mechanisms and challenges.

As a part of this assignment, your job is to read the paper, and write one-page summary (Single-space, 12 pt. font size) on the key points that are discussed in this paper. In addition, you need to write your views on this topic.

Hint: To write the views, you can use concepts that are discussed in the paper, in the 5410 class lectures, and the limitations of the published material that you discovered while reading the paper.

Part B - Submission requirement:

A PDF file with the summary of the paper, and your views. You need to cite all the materials that you have referred including the given paper in your PDF file. Do not forget to add in-line citation