DATA CENTER SCALE COMPUTING - LAB 5

Objective - This lab is designed to help you learn and implement Apache kafka using confluent. By the end of this lab you will be able to

- Connect with confluent and create a kafka cluster.
- Create topics, producers and consumers
- Understand the partition system

Instructions

- You are required to implement producers and consumers in python
- There will be an interview grading for this lab. You should be able to logically explain the reason behind the behavior.

Task - 1 [25 points]

1. Use the json file provided in the repo as the data source
2. You should produce a key value pair with key to be the unique identifier
3. Create a topic named - “customerOrders”
4. Create a producer and a consumer with a group.id = ‘id-1’.
5. Print the consumed messages with the partition and the offset

Task - 2 [25 points]

1. Create a consumer for task 1 with group.id =’id-1’.
2. Run the consumer file
3. Print the consumed messages with the partition and the offset
4. Explain your observations

Task -3 [25 points]

1. Create a consumer with a different group.id of your choice
2. Run the consumer file
3. Print the consumed messages with the partition and the offset
4. Explain your observations

Task - 4 [25 points]

1. Pick a datasource of your choice - json/csv/any d
2. Create a topic

3. Create a producer and a consumer
4. Run the producer you created and the producer from Task - 1 and publish the messages to their respective topics. Use the same consumer created in task 1 to consume the the messages from the two topics
5. Display results and write your observations