Pipelines Assignment

MSc in Business Analytics

Big Data Systems

Deadline: Wed April 13th 2022, 23:59

Task 1 [50 points]

Use Airflow to define a small workflow that leverages BashOperator that contains the following tasks:

- Task 1.1: Creates a string variable called "firstName" assigning to it the string that corresponds to your first name with all letters in lowercase.
- Task 1.2: Creates a string variable called "lastName" assigning to it the string that corresponds to you last name with all letters in lowercase.
- Task 2.1: It gets the firstName as input and outputs the same string with the first letter capitalized.
- Task 2.2: It gets the lastName as input and outputs the same string with the first letter capitalized.
- Task 3: Displays on screen a string that concatenates the firstName and the lastName adding a space character in between.

Your deliverable should be a Python file that contains the respective DAG definition (named "id-t1.py", where "id" refers to your student id).

Task 2 [50 points]

Use the kafka-python package to create a KafkaProducer to publish messages in a topic called "clima". Each message should be in JSON format, containing one value for a temperature and one for a humidity reading (both integers, you can select names of your choice for the keys). Each message should also contain a timestamp representing the time point during which the measurements took place. Moreover, use the send() function to publish five example messages (with values of your choice). Also create a KafkaConsumer that subscribes to the respective topic and prints the whole history of messages on screen. Finally, make sure that your code is provided to a Kafka deployment hosted locally (the port 9092 is used). Your deliverable should be a Python file that contains your code (named "id-t2.py", where "id" refers to your student id).

Submission instructions & honor code

Your code files should be fully replicable and readable (documentation comments are required and appreciated). Your code should be ready to be executed (e.g., containing all the required import statements). Code failing to execute or producing wrong results will be penalised. You understand that this is an individual assignment, and as such you must carry it out alone. You may discuss with your fellow students to better understand the tasks/questions but you should not ask them to share their answers with you or to help you by giving you specific advice.

GOOD LUCK!