POLITECNICO
MILANO 1863

Apache Kafka

Alessandro Margara
alessandro.margara@polimi.it
https://margara.faculty.polimi.it

Rules

- Complete the java files as indicated by each exercise
  - Only add code in the locations indicated by comments
- Complete the README.md file with
  - Number of your group (from the group registration document)
  - Name of each group member
  - Number of partitions needed for each exercise (minimum, maximum)
    - Maximum can be set to N if no upper bound is required
  - Number of consumers needed for each exercise
    - Same as above
  - Arguments to start each consumer
- Create and submit a single zip file with your project
  - Name of the file: kafka-groupXX.zip (XX is the number of your group in the group registration document)
  - Submitted from the contact email provided in the group registration document

Assumptions

- One instance of the Producer class publishes messages into the topic "inputTopic"
  - The producer is idempotent
  - Message keys are String
  - Message values are Integer
- You may set the number of partitions for "inputTopic" using the TopicManager class
  - Indicate in the README.md file the minimum and maximum number of partitions allowed
- Consumers take in input at least one argument
  - The first one is the consumer group
  - You may add any other argument you need

Exercise 1

- Implement a FilterForwarder
  - It consumes messages from “InputTopic”
  - It forwards messages to “OutputTopic”
  - It forwards only messages with a value greater than threshold (which is an attribute of the class)
  - It provides exactly once semantics
    - All messages that overcome the threshold need to be forwarded to once and only once

Exercise 2

- Implement a AverageConsumer
  - It computes the average value across all keys (i.e., the sum of the last value received for all keys divided by the number of keys)
  - It prints the average value every time it changes
  - It does NOT provide guarantees in the case of failures
    - Input messages may be lost or considered more than once in the case of failures
    - The average value may be temporarily incorrect in the case of failures