Project #2

Assignment #1 – Data Transport with Apache Kafka

14th November 2024

Advanced Infrastructures for Data Science

Pedro Neves, pedroneves@dei.uc.pt, 2024/2025

Master in Data Science and Engineering (MDSE) Course

Task #1 – Producer and Consumer with 1 Topic 1 Partition

Assignment #1

- Objective
  - Introduce the basics of Kafka producers and consumers by working with a single topic and partition

- Instructions
  - Create a Kafka topic called **task1-topic** with one partition
  - Write a Kafka producer, which sends a random temperature reading (e.g., {"sensor_id": 1, "temperature": 22.5}) every 2 seconds to **task1-topic**
  - Write a Kafka consumer that subscribes to **task1-topic**, reads each temperature message, and prints it to the console

Task #2 – Multi-Partition Topic with Multiple Consumers and Consumer Groups

Assignment #1

- Objective
  - Explore the effect of multiple partitions, consumers, and consumer groups on message distribution

- Instructions
  - Create a new topic called **task2-topic** with 3 partitions
  - Write a Kafka producer to send user activity logs (e.g., {"user_id": "user1", "activity": "login"}) to **task2-topic**
  - Create a consumer group called **activity-group** and set up four consumers within this group
  - Observe and document the behavior when:
    - Three consumers in **activity-group** are active
    - Two consumers in **activity-group** are active
    - Four consumers in **activity-group** are active

Task #3 – Multi-Topic Producer and Consumer with Consumer Groups

Assignment #1

- Objective
  - Explore the effect of working with multiple topics, consumer groups, and stream processing concepts

- Instructions
  - Create two topics: **purchase-topic** and **user-activity-topic**
  - Write a Kafka producer that sends purchase transactions to **purchase-topic** (e.g., {"user_id": "user2", "amount": 15.0, "item": "book"}) and user actions (e.g., {"user_id": "user2", "activity": "page_view"}) to **user-activity-topic**
  - Create two consumer groups:
    - **purchase-group** for processing messages in **purchase-topic**
    - **activity-group** for processing messages in **user-activity-topic**
  - Write a consumer (**purchase-group**) that aggregates the total amount spent by each user and prints a running total for each user
  - Write a consumer (**activity-group**) that counts the activities per user and prints the activity count

Good
Work