

Exercise 8: Stream Processing with Kafka Streams (Part II)

Deadline: 02.06.2024; 23:59 CEST

Project Abstract

In the ninth week's exercise on Event-driven and Process-oriented Architectures we will investigate the use of Kafka Streams for stateful stream processing.

① Labs

In the practical part of the lecture we ask you to work on the following lab to get started with stateful processing in Kafka Streams:

- Lab13Part1 – LeaderboardApp¹: This lab shows the Kafka Streams-based implementation of the Leaderboard example for stateful stream processing introduced in the lecture. Here you find a topology involving streams, tables and stateful operations like joins, groupings and aggregations. Try to run the project and study the implementation of the stream processing app.
- Lab13Part2 – Eye Tracking Events Processing (Part2)²: This lab further demonstrates the Kafka Streams-based implementation in the eye-tracking domain. The project covers both stateless (content filter, event filter, event translation, event routing) and stateful processing of events (grouping, aggregation, joining). It consists of four modules: a fixation producer, a click stream producer, an event processing topology (with interactive querying and monitoring), and a fixation consumer. Try to run the project, study the implementation of the event processing topology and extend it as described in the *Exercise* Section of the project README.

② Suggestions for Software Project

For your software application in Assignment 2 we suggest that you implement one or several stream processing applications including both stateless and stateful stream processing:

- At least several stateless operations (Week 8)
- Usage of both streams and tables (Week 9)
- Consider data from more than one stream (Week 9)
- Interactive queries (Week 9)
- Windowed operations (Week 10)

We suggest to use Avro schemes for serialization and deserialization of records in Kafka (registry-less or with a schema registry).

Software Project Expectations

The labs in Part 1 of this sheet are meant to be done during class and are **not graded**. In Part 2 we expect you to prepare a report on the project-related implementations that you have done in your group. For each developed stream processing application the **topology** should be described in **graphical**

¹<https://github.com/scs-edpo/lab13Part1-kafka-streams-leaderboard>

²<https://github.com/scs-edpo/lab13Part2-kafka-streams-eyeTracking2>

form and you should **elaborate** on the chosen **stream processing patterns**. Feel free to use the provided data sets for your experiments.

Hand-in Instructions

The report and implementation for Part 2 will **be graded** as part of the software application for your group to be handed in for the second part of the lecture until the semester end by the deadline indicated on the top of this sheet. Each group member **must explicitly indicate** which part she/he/they has/have been working on. Please approach the tutors for individual feedback regarding this assignment.