

Title of the course: Data Streaming

Instructor: Martina Šestak (University of Maribor, Slovenia)

Aims of the course: Students will learn about Data Streaming concepts and how to combine different processing approaches when designing big data architectures.

Learning objectives: After completing the course, students should be able to differentiate between batch and streams processing paradigms, analyse and understand the appropriate application scenarios for each, and design and implement big data pipelines to integrate different data technologies.

1. Data Streaming: understand the fundamental concepts of Data Streaming and Streams Processing.
2. Batch vs Streams Processing: understand the difference between two processing approaches and when each is applicable.
3. Kafka platform: learn about the features, capabilities, and architecture of Kafka and available components and libraries.
4. Big data architectures: learn about lambda and kappa architectures, layers of big data architectures and how to build data pipelines.

Structure: In the first part, the fundamental topics about data streaming and Kafka will be presented. The second part involves solving practical assignments in groups of students, where several case studies will be given. The students will need to analyse processing needs, and design and implement data pipelines that integrate different technologies.

Topics: Data Streaming, batch and streams processing, Kafka, lambda and kappa architecture.

Assignments: Assignments based on case studies and practical exercises to be solved in international teams of 2 to 3 students.

Technical Prerequisites: Students will be provided with a *docker-compose* file, which includes the configuration for setting up the Kafka environment and the tools necessary to complete the exercises. This requires a local installation of the Docker environment (including Docker Desktop for easier container management).