

DTC-MBD-525 Stream Processing

SEMESTER: Fall
CREDITS: 3 ECTS
LANGUAGE: Spanish/English
DEGREES: MBD

Course overview

The purpose of this course is to provide students with a fundamental understanding and an extensive practical experience on how to process and analyze data in a streaming fashion with the most important tools in the big data ecosystem.

By the end of the course, students will:

- Understand the basic principles of the stream processing.
- Have practical experience with the most important tools in the stream processing (Spark Streaming, Spark Structured Streaming and Kafka Streams).
- Have well-formed criteria to choose the most appropriate tool for a given streaming application.

Prerequisites

Students willing to take this course should be familiar with Scala and Java programming languages. It will also be desirable to know the basic concepts of the streaming framework Kafka.

Course contents

Theory

1. Introduction
 - 1.1. ¿What is Stream Processing?
 - 1.2. Tools of Stream Processing: Samza, Storm, Spark Streaming, Spark Structured Streaming, Kafka Streams and Flink.
 - 1.3. Kafka revisited.

2. Spark Streaming
 - 2.1. Introduction
 - 2.2. DStream and RDD processing. Kafka integration.
 - 2.3. Stateless (transformations, filtering) and stateful operations (aggregations, joins)
 - 2.4. Windowing.
3. Spark Structured Streaming
 - 3.1. Introduction
 - 3.2. Input table and result table. Kafka integration.
 - 3.3. Stateless (transformations, filtering) and stateful operations (aggregations, joins)
 - 3.4. Windowing. Late data handling. Watermarking.
4. Kafka Streams
 - 4.1. Introduction
 - 4.2. KTables and KStreams. Stream – Table duality
 - 4.3. Stateless and stateful operations. Local state stores. Streams DSL.
 - 4.4. Windowing. Interactive queries. KSQL

Practice

Each session will be split into two parts: an introduction of the theoretical concepts and a lab practice using programming languages (Scala and Java) to process data. The lab will be resolved by the teacher at the end of the session.

Textbook

- Notes prepared by the lecturer (available in Moodle)

Grading

- **Evaluation exercise at the end of the course.**
 - Theoretical examination (test): 50%
 - Practical exercise : 50%