Computational Data Science Program

Course outline

Big Data Analytics and Visualization (5 ECTS)

Core Concepts of Big Data

- What is Big Data?
- Evolution of Big Data
- Why Big Data?
- Characteristics of Big Data
- Applications of Big Data

Big Data Analytics

- What is big data analytics?
- Data analytics: Key concepts
- Descriptive, Diagnostic, Predictive and Prescriptive Analytics
- Data Warehouse Architecture
- Technologies used in Big Data analytics

Big Data with Hadoop

- The core modules of Hadoop
- Hadoop MapReduce
- Hadoop YARN

Big Data with Spark

- Spark jobs and APIs
- Resilient Distributed Datasets (RDD)
- PySpark with DataFrames
- Prepare Data for Modeling using PySpark

Big Data Visualization for Decision Making

- Manipulating geographic data using GeoPandas

References

- Tomasz Drabas and Denny Lee, Learning PySpark,2017
- Ofer Mendelevitch, Casey Stella, Douglas Eadline: Practical Data Science with Hadoop and Spark, 2017
- Arshdeep Bahga andVijay Madisetti: Big Data Analytics: A Hands-On Approach, 2019
- Mahmoud Parsian: Data Algorithms with Spark Recipes and Design Patterns for Scaling Up Using PySpark, 2022