Introduction to Big Data & Spark

(Laurent Lecornu)

This course introduces engineering students to the world of Big Data, focusing on distributed storage and computing concepts. It provides hands-on experience with the Spark library, enabling students to execute basic commands in Spark environments. The knowledge gained is crucial for engineers working in data-intensive industries and large-scale data processing applications.

Key concepts covered:

- Fundamentals of Big Data and its challenges
- Distributed storage and computing principles
- Apache Spark architecture and ecosystem
- Basic Spark operations and transformations
- Data processing and analysis using Spark
- Real-world applications of Big Data technologies

By the end of this course, students will be able to:

- [BC-03] Understand and explain the principles of distributed computing in the context of Big Data
- [BC-04] Design and implement simple data processing pipelines using Spark
- [BC-07] Execute commands and operations in Spark environments

Prerequisites :

- Being familiar and efficient with Python programming