

Advanced Big Data Architecture

(Laurent Lecornu)

This course delves into advanced concepts of Big Data architecture, focusing on the Hadoop ecosystem and its associated technologies. It provides in-depth knowledge of key components and emphasizes both theoretical aspects and practical implementation of Big Data cluster architecture.

Key concepts covered:

- Hadoop ecosystem and its core components
- Advanced technologies: Cassandra, Kafka, and Spark Streaming
- Big Data cluster architecture design
- Functional programming for horizontal parallelization
- Installation and configuration of Big Data modules
- Creation of processing pipelines using Cassandra, Kafka, and Spark Streaming

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

- [BC-01] Configure and manage Big Data clusters for optimal performance and scalability
- [BC-04] Design and implement complex Big Data architectures using Hadoop ecosystem technologies
- [BC-07] Create and optimize processing pipelines for real-time and batch data processing

Prerequisites :

- Proficiency in Python programming
- Basic knowledge of Linux commands