

Title: *Distributed Technologies for Big Data Processing and Blockchain*

Teacher: Luca Grilli and Fabrizio Montecchiani

Contact: luca.grilli@unipg.it; fabrizio.montecchiani@unipg.it

Indicative period: January - May 2022. (interested students should contact the teacher as soon as possible in order to define the schedule of the lessons).

ABSTRACT. Both Big Data processing systems and blockchain technologies have a distributed nature. Big Data encompasses large volumes of complex and dynamic data, which are often characterized by a multimodal nature and by a high generation rate. As a consequence processing and managing Big Data pose challenges that go beyond the capabilities of conventional software and systems. This course presents some of the main distributed computing paradigms and technologies that can be exploited to process Big Data. Blockchain technologies also rely on distributed systems, but this is required to improve data security and to avoid trusted central entities, rather than to expand data processing capabilities. A blockchain is a chain of immutable transaction blocks that are memorized across multiple networked data stores - the blockchain network - each of which can be owned and handled by autonomous organizations with conflicting interests. Blockchains make strong use of cryptography and of (distributed) trustless consensus mechanisms to guarantee irreversibility, authenticity and integrity of their content. This course gives an overview of the main types of blockchain technologies, including programmable blockchains and smart contracts.

PROGRAM

- Introduction to Big Data Processing.
- The MapReduce data processing model, the Apache Hadoop platform, and the Spark processing engine.
- Fundamental concepts of information security and cryptography.
- Introduction to blockchain technologies, blockchain components, and types of blockchains.
- From Bitcoin's blockchain to programmable blockchains, the Ethereum blockchain, and the Solidity language for smart contracts.