

Distributed architectures for big data processing and analytics

Exam Rules – Academic Year 2019-2020

The exam consists of a written exam.

Only students regularly registered for the exam through the Teaching Portal (Portale della Didattica) are admitted to the classroom for the written exam.

- Students must provide their own identity document to be allowed to take the exam
- The exam lasts 2 hours
- Paper books and paper notes are allowed. Instead, no electronic devices (PC, laptop mobile phone, calculators, etc.) are allowed.

The written exam includes

- 2 programming exercises (max 27 points)
 - Design and develop programs based on the MapReduce programming paradigm and Spark APIs
- 2 questions/theoretical exercises (max 4 points)
 - Topics of the questions/theoretical exercises
 - Technological characteristics and architecture of Hadoop and Spark
 - HDFS
 - MapReduce programming paradigm
 - Spark RDDs, transformations and actions
 - Spark SQL
 - Data mining and Machine learning libraries for Big data (Spark MLlib, GraphX/GraphFrame)
 - Data streaming analytics (Spark Streaming, Flink, Storm, Kafka, ..)
 - Relational and NoSQL databases for big data (HIVE, HBase)

The final exam grade is given by the sum of the grades of the two parts.

The evaluation of the programming exercises is based on the correctness and efficiency of the proposed solutions.

To pass the exam the following constraint must be satisfied:

- Final exam grade ≥ 18