

Distributed architectures for big data processing and analytics

Exam Rules – Academic Year 2019-2020 – Online exam

The exam consists of an **online** written exam. The online written exam will be conducted through the Polito's **Exam** tool combined with the **Lockdown browser + Respondus** proctoring system.

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

- Students have to provide an identity document including a photograph (Photo ID) to take the exam
- The exam lasts 2 hours
- Each student needs a **PC compatible with the Respondus (Lockdown browser) system** and an **adequate internet connection**.
- Each student must take the exam **in a room alone. Only the PC used to take the exam must be on the desk. No other objects** can be present **on the desk**.
 - **Books, notes, empty sheets and any other paper materials are NOT allowed.**
 - **Electronic devices** of any kind (PC, laptop mobile phone, calculators, etc.), apart from the PC used to take the test, **are NOT allowed.**
 - Exceptionally, to support the students during the exam, the exam questions will include a link to the PDF version of the slides of the course. Students can use that link to access the PDF version of the slides of the course inside the Lockdown browser. **The availability of the linked folder for the entire duration of the exam cannot be guaranteed. Hence, the exam is valid also if, for any reason, the provided linked folder is not available or accessible during the exam.**
- If any problem preventing the continuous remote supervision of the test arises, **the guidelines approved by the University will be applied.** Such problems include, but are not limited to:
 - Hardware PC problems
 - Lack of audio from the microphone or video from the webcam (even temporary)
 - Lack of network connection (even temporary)

On teacher discretion, some students may be required to take an oral interview in order verify that the written exam has been taken following the exam rules.

The online written exam includes

- 2 programming exercises (max 27 points) - You will write your code in the text box of the open questions associated with the problem specifications.
 - Design and develop programs based on the MapReduce programming paradigm and Spark APIs
- 2 multiple-choice questions/theoretical exercises (max 4 points)
 - Topics of the multiple-choice 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