Project #2 – Big Data Processing Techniques

Submission Instructions

12th December 2023

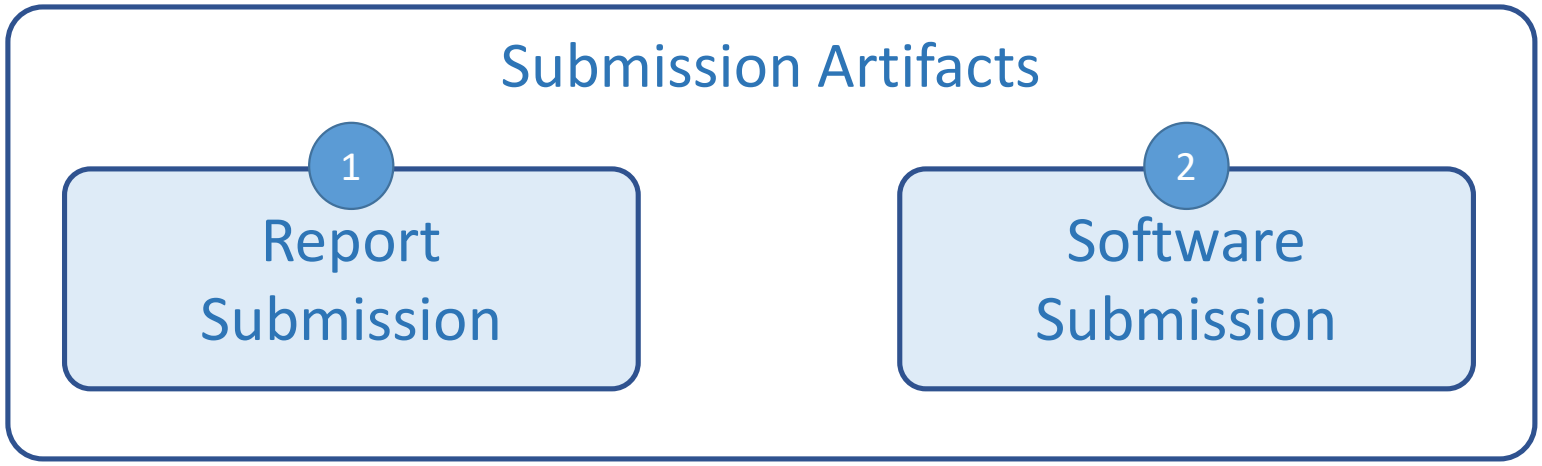
Advanced Infrastructures for Data Science

Pedro Neves, pedroneves@dei.uc.pt, 2023/2024

Master in Data Science and Engineering (MDSE) Course

Project #2: Submission Instructions

- Submission must be made in **Inforestudante**
- Deadline: **12th Dec**, 23h59
- Submission is composed of a "**Report**" and "**Software**"

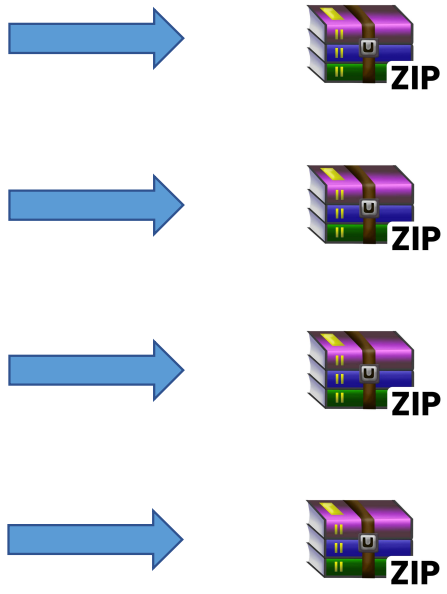


Project #2: Software Submission

- Project #2 – Software to be Submitted

Submit 4 zip files

1. Data Processing with Hadoop MapReduce (Assign #1)
   MapReduce scripts
2. Data Processing with Spark RDD (Assign #2)
   pySpark RDD scripts
3. Data Processing with Spark SQL (Assign #3)
   pySpark SQL scripts
4. Data Processing with Spark Streaming (Assign #4)
   pySpark Streaming script



Project #2: Report Submission

- Project #2 – Report to be Submitted
  - Briefly list (with bullets) the **sequence of steps** that are required to ingest and transform the data for each assignment/task objective
  - Include the **output** obtained for each assignment/task
  - The report size should not exceed **6 pages**
  - Should be delivered in **.pdf** format

Submit 1 pdf file

Good
Work