Project

Guidelines and deadline

You are asked to form groups of 2/3 students max.

Each group has to write a report including:

1. a description of the adopted solution 4 points

2. designed algorithms plus related global comments/description 4 points; comments to main fragments of code 3 points

3. experimental analysis, concerning in particular scalability 3 points

4. comments about the experimental analysis outlining weak and strong points of the algorithms. 3 points

5. an appendix including all the code. 2 points

A pdf version of the report has to be sent via email (dario.colazzo@dauphine.fr) by March 1st, 2026.

Finding connected components in graph

- The algorithm is described in this *short* paper
  - https://www.cse.unr.edu/~hkardes/pdfs/ccf.pdf
- The work to consists of understanding the 2 MapReduce algorithm, and coding it into Spark by using RDD
- Experimental analysis on graphs of increasing size, comparing the two implementations also.
- You can obtain the graphs staring from the references in the paper. You can also generate synthetic graphs, of different/ increasing size.