Project

- The project is about finding connected components of large graphs.

- You will find here a paper describing the pseudo-code of a MapReduce algorithm

    https://www.cse.unr.edu/~hkardes/pdfs/ccf.pdf

- You will also find illustrations describing the algorithm

- The work to do consists of implementing the two proposed variants in PySpark by means of RDD.

- Write a technical report, around 20 pages max, with main points about implementations and experiments on reasonably large graphs. *Main parts* of the code need to be commented.

- Groups of 2/3 person max are allowed.

- Send the report + the code by email by the indicated deadline.

    · ***Deadline February 25 , 2024***