# Project

# Finding connected components in graph

- The algorithm is described in this paper
  - https://www.cse.unr.edu/~hkardes/pdfs/ccf.pdf
- The work to consists of understanding the MapReduce algorithm, and coding it into Spark by using both RDD and DataFrames
- Python implementations must be provided
- Experimental analysis comparing the RDD and DataFrame versions has to be conducted on graphs of increasing size
- For small graphs use Databricks, for bigger ones use the cluster

# Guidelines

The report should contain

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 4 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 the code. 2 points