

## Homework n° 2 – Apache Spark

The second homework concerns the development of a program based on Apache Spark that combines the information about infected people and vaccines in order to:

- 1) Provide some insight about COVID-19 pandemia (analyzing data available at the following link (<https://github.com/pcm-dpc/COVID-19>);
- 2) Analyze the effects of the vaccine on the spread of the epidemic for each region (analyzing both the following links (<https://github.com/pcm-dpc/COVID-19> and <https://github.com/italia/covid19-opendata-vaccini>))

The output must be reported within a pdf file of 8/10 pages (whose name will be name\_surname\_homework\_2), whose structure should include the following information:

- Problem introduction and dataset characterization;
- Description of developed code with the obtained results and analysis.

This homework can be done individually or in groups of up to two people. In the case of groups, the output must be at least 15 pages.

The deadline for the homework is Friday, May 7, 2021.