

## Homework no 3 – Apache Spark

The third homework concerns the development of a program based on Apache Spark for extracting some useful information from the YELP dataset (<https://www.yelp.com/dataset>).

The output of the analysis must be reported within a pdf file of 8/10 pages (whose name will be name\_surname\_homework\_3), whose structure should include the following information:

- Problem introduction and dataset characterization;
- Processing queries and algorithms based on Apache Spark with the related discussion

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 10 pages.

The deadline for this homework is Monday May 30, 2022.