

Homework nº 2 – Hadoop PIG

The second homework concerns the development of a program based on Apache PIG for extracting some useful information from the review file of 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_2), whose structure should include the following information:

- Problem introduction and dataset characterization;
- Processing queries and algorithms based on Apache Pig 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 16, 2022.