# Challenges

The challenges for the HW and SW Architecture for Bog Data course must be organized in the following 3 chapters:

1. Challenge description
2. Description of the proposed methodology and architecture
3. Experimental results

Furthermore, the chosen challenge must be addressed respecting the following two constraints:

1. Using Spark with Python.
2. Using at least one NoSQL DB.

Each group must present during the exam date:

1. The developed code;
2. Project description in PDF format.

In the following we describe the proposed challenges.

## Recommendation

The aim is to define a recommendation algorithm using the MovieLens dataset information to suggest movies to users of the site.

The proposed approach will be evaluated in terms of MSE, RMSE and accuracy.

Dataset: MovieLens 20M (<https://grouplens.org/datasets/movielens/>).

## Community detection

The aim is to define a community detection algorithm to unveil users’ subset sharing same taste and preferences.

The proposed approach will be evaluated in terms of NMI, ARI and Accuracy.

Dataset: DBLP (<https://snap.stanford.edu/data/com-DBLP.html>).

## Natural Language Processing & Sentiment Analysis

The aim is to define algorithm for evaluating user’s reviews (positive or negative) or for analyzing user’s sentiment about a given business object.

The proposed approach will be evaluated comparing the obtained result with review stars.

Dataset: Yelp (<https://www.yelp.com/dataset/challenge>)