# PDP Assignment 1

Samuel Kuik,

580221,

20-06-2022

<https://github.com/Zalmwel/PDP>

## Explanation

1A

(in order of usage)

We start with ‘steps’, to run our functions in one go, giving the output of the previous function to the next function.

Graphical user interface, text, application

Description automatically generated

The ‘mapper\_get\_ratings’ function creates a key/value pair of movieID and 1 , with movieID being the key and 1 (or value) being the amount of ratings per row read.

A picture containing graphical user interface

Description automatically generated

The reducer counts when a movie received a rating and sums these values up for each movieID.

Text

Description automatically generated with low confidence

Finally, the reducer\_sorted\_output function iterates over each id in movieID and shows its number of ratings.

## Text Description automatically generated with medium confidence

1B

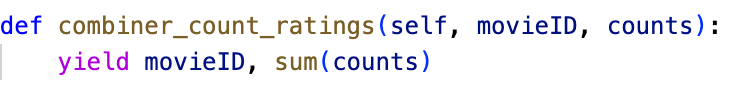
To sort the ratings from movies with the most ratings to least requires a few additional steps.

We start off the same way as in 1A, first the mapper function creates a key/value pair.

Text

Description automatically generated

Next, instead of immediately having the reducer making the sums for each movieID, we first use the combiner to aggregate the values, reducing the number of rows greatly.



After this, the reducer makes a sum for each movieID and the reducer\_sorted\_output again iterates over each id in movieID and shows its number of ratings.

Text

Description automatically generated

## Results

Assignment 1A: Assignment 1B:

