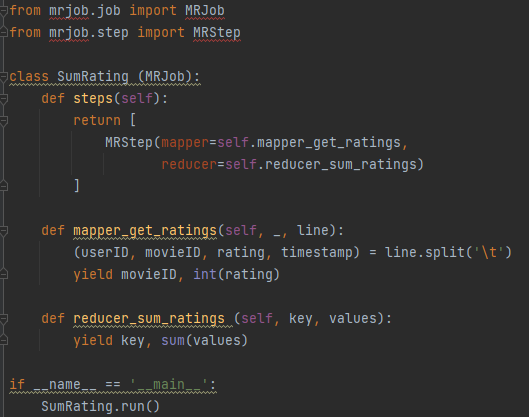
Abdullah Ghayumi

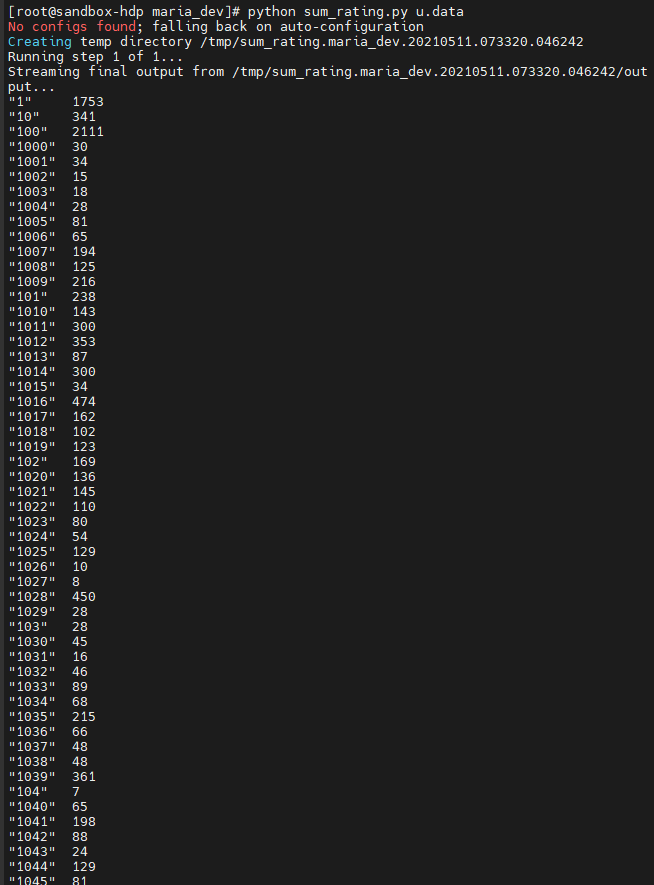
634072

Github URL: <https://github.com/Aras53/Hadoop>

After struggling to get everything installed, I finally got it working. Afterwards I used the default script that was presented during the class to create the sum ratings script for each movie.



The mapper\_get\_ratings function will get the various columns and values from the data file. Then it will only yield the movieID and the rating that I had casted to int. I casted it to int to be able to use it in the reducer\_sum\_ratings function. Here I take the key, that is the movieID and sum all the ratings that the movie got. With that done, I used the following command `python sum\_rating.py u.data` in the terminal to get the following results:



For the sorting assignment I used a chaining of MRSteps. First, I got a list with movieID and the total ratings they got. With that list, I created a new list on values that adds all the keys and values in the values list. Besides that, I made the total ratings a string so that the sorting would work, with ints it gave a result like this: 1, 10, 100 etc.

I tried doing this with sorted() and .sort(), but was not successful so this was my final solution.



The result is a list with the movieID on the left and the number of ratings sorted on the right in string format.

