Hadoop - Assignment 1

GUIDO VAN DIJK   
IT2B  
648539

GitHub Link: <https://github.com/guidovdijk/hadoop-assesment-1>

12/05/2022

# Steps to take to execute:

The steps assume that the user already has the ml-100k folder on his machine and the python file will be added on the same path as the folder.

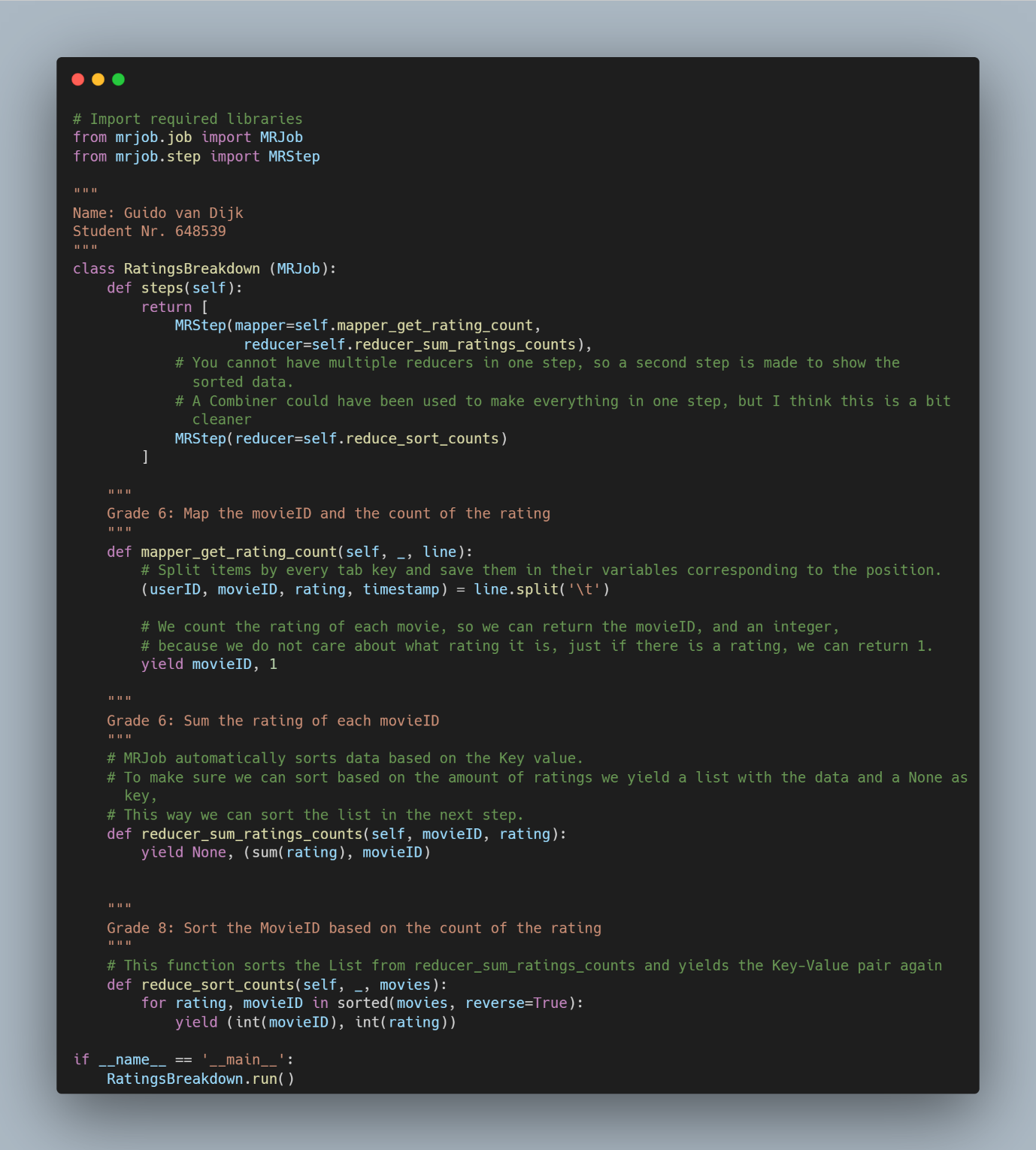
Structure  
| …  
| - <name>.py  
| - ml-100k   
| - u.data

If not see the presentation of the first week on how to add it.

1. Start the Virtual Environment
2. Start PuTTy
3. Login as root user – **su root**
4. Create a new python file in the root **touch <name>.py**
5. Open the file **vi <name>.py**
6. Copy the code from the repository
7. Paste it in the python file “go into insert mode **i** > ”
8. Save and quite **!wq**
9. Run the python file: **<name>.py -r hadoop --hadoop-streaming-jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar ./ml-100k/u.data**

# Steps of the code

Steps in the code are shown in the file itself. Can also be read in the repository: <https://github.com/guidovdijk/hadoop-assesment-1/blob/main/main.py>



# Result of the code

Shape, rectangle

Description automatically generated