Assignment 1 – Parallel Distributed Processing

Student: Jaap van Dalen

Student number: 630564

Github-link: <https://github.com/JaapVanDalen/Inholland-P4.4-PDP-Assignments>

# Requirements

Have python installed.

Make sure you have pip installed.

Make sure you have MrJob installed. (pip install mrjob)

# Steps to execution

# Download the movielens dataset from Inholland. Then you can move to the on the folder where you have the file rating\_per\_movie.py and u.data stored. then run the command: py rating\_per\_movie.py u.data

# Code Explanation

There are comments above each line of code. These will explain every single line of code in the script.

# GitHub explanation

Sum\_rating\_per\_movie.pig code for a 6

Sum\_rating\_per\_movie\_sorted.pig extra: for an 8

documentation.docx this document

# result

first screenshot assignment 1 for a six, second screenshot is the output of the script `sum\_rating\_per\_movie\_sorted.py`

 