Getting the required files:

wget [https://raw.githubusercontent.com/ash...](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbTM2bTF2aS00R3hzMVhoSFpzbmxXS0NOUEZJQXxBQ3Jtc0tuUWRabzlNVVoxVUoxeDNWTVNlaUY1Q2NqOUg0OEEybnFyT2g3cHhFdHVOOWpIUURYQmJiUWVZMHNsT2RCRmRnU01yZ3hmeHN4dVlFa2dxRURINmRxczFfMDFPV3BRWEJqdnV4cUsxUG5WZXZRZ2dOcw&q=https%3A%2F%2Fraw.githubusercontent.com%2Fashaypatil11%2Fhadoop%2Fmain%2Fratings.data&v=td8tcX9OlBk)

wget [https://raw.githubusercontent.com/ash...](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbl9iSkhEWG5tX2dhVDQyQVBOWlBuTDd5b01HUXxBQ3Jtc0tuQmljdmRiU3RkWnpqMC1aSGZkQmN2dmdWY3NENXlCMHBTOERTYzlaZlFiSFJwRUZ0dU5PM01ZQlVxRnEybkJ4RmhraGNldElKcnNQcGRqYS10NGxwOEFLbHlrVEZFWjlGZTdyS1lyUGVzSXgxeGZzVQ&q=https%3A%2F%2Fraw.githubusercontent.com%2Fashaypatil11%2Fhadoop%2Fmain%2FMovieRatings.py&v=td8tcX9OlBk)

Running locally: python MovieRatings.py ratings.data

Running on Hadoop cluster:

python MovieRatings.py –r hadoop --hadoop-streaming-jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar ratings.data

Installing mrjob on HDP 2.6.5 (be sure to "su root" first, as shown in the video.) yum-config-manager --save --setopt=HDP-SOLR-2.6-100.skip\_if\_unavailable=true yum install [https://repo.ius.io/ius-release-el7.rpm](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbWlLeVNtYkMyanprT1hlSWJoeWJCMVF1Y0FmUXxBQ3Jtc0ttdkktSGJMVncyaHkzbVJnM1lyVmt2eVhiVUMyWGJSWkFkaTBtdXA0bEo2LVl3SXpWeHlfTHluR203Z183VFRhdVVyY1YwMlQ4NlhjZXhETENaZXZiRzlKREdrSkh1Nms0QnRRYlB1MHhkOExUcUliaw&q=https%3A%2F%2Frepo.ius.io%2Fius-release-el7.rpm&v=9KyqEy1bWWA) [https://dl.fedoraproject.org/pub/epel...](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbnB0U2s0eG9QY3diOVFHUVhPcDZHY1o1SFlTd3xBQ3Jtc0ttU3B0ZWZ4M29vRkppWmxqcVJsTTNUOEpnWk9LUXJvc0pxYTJUUWdIZzRSRlYtWk9xUDMwSjZSV05GcDR3WHhXcElablVPeGpFSU1zWmNucjVwdl9URk9hUkhzSmk2UXU1WURhUE1hbU1UdHpQOTUtYw&q=https%3A%2F%2Fdl.fedoraproject.org%2Fpub%2Fepel%2Fepel-release-latest-7.noarch.rpm&v=9KyqEy1bWWA) yum install python-pip pip install pathlib pip install mrjob==0.7.4 pip install PyYAML==5.4.1 yum install nano