I found a pretty good tutorial at [http://www.rohitmenon.com/index.php/apache-pig-tutorial-part-1/.](http://www.rohitmenon.com/index.php/apache-pig-tutorial-part-1/) You get the data from

[hadoop@pig ~]$ wget https://github.com/rohitsden/pig-tutorial/blob/master/movies\_data.csv?raw=true -O movies\_data.csv

I did have to make some modifications. Menon never said, but he did his tutorial in local mode. I wanted to do it in mapreduce, just because. I had to copy my movie data into an HDFS directory "wk5."

[hadoop@pig ~]$ hadoop fs -copyFromLocal ~/movies\_data.csv wk5

That made my LOAD command different.

grunt> movies = LOAD 'wk5/movies\_data.csv' USING PigStorage(',') as (id,name,year,rating,duration);