

Exercise 23 - PIG

When you open your cloudera image make sure to start the manager first and ensure that the namenode leaves "savemode". Start HUE from the first page and go through the setup process, all examples and add ons need to be installed before you can use PIG.

Go to this page:

<https://github.com/rohitsden/pig-tutorial>

and download the example file containing the movies_data.csv. This file comes from the following tutorial which also wraps up what we did today:

<http://www.rohitmenon.com/index.php/apache-pig-tutorial-part-1/>

Once you have unpacked movies_data.csv you need to upload this one to HDFS, either use the command line or the file manager for that.

The only thing to do for today's exercise is to filter the movies file for movies made after 1990. Please write a pig script that does that and which saves the results into a file. This file needs to be transferred to the regular filesystem. Please use the command "less" on the file and send in a screen capture of the first couple of lines of that file.