Pig Installation Guide - 0.13.0

NOTE: If a dedicated Hadoop user is present (hduser), Pig needs to be installed in that user account. However, Pig can run standalone as well.

Download and Install Pig:
-- Switch to appropriate user where Hadoop is installed or skip this step: $su hduser
-- Go to the home directory: $cd ~
-- Download Pig using the command:
$curl -o pig-0.13.0.tar.gz http://www.eu.apache.org/dist/pig/pig-0.13.0/pig-0.13.0.tar.gz
-- If the above link fails to work, download Pig from the apache website (tar.gz file).
-- Unzip the folder with the command: $tar -xvzf pig-0.13.0.tar
-- The above command will create a new Pig file (pig-0.13.0), to check use the command: $ls -l
-- Rename the unzipped folder to "pig": $mv pig-0.13.0 pig
-- Check the present working director: $pwd #should be /Users/Username
-- Move the Pig folder from the pwd to /usr/local: $sudo mv /home/hduser/pig /usr/local/pig
-- Change owner of Pig folder in /usr/local to the Hadoop user if required (hduser in this case): $sudo chown -R hduser:hadoop /usr/local/pig
-- To check permissions of the Pig folder, go to /usr/local and enter the command: $ls -al

Setup Pig Configuration Files:
The .bash_profile file has to be configured in order to complete the Pig Setup.
-- Switch to the Hadoop user where pig is installed (hduser in this case) or skip this step: $su hduser
-- Go to the home directory inside hduser: $cd ~
-- Check if the .bash_profile file exists: $ls -al
-- Open the .bash_profile file using TextEdit: $open -a TextEdit .bash_profile
-- Append the following code to the end of the .bash_profile file:

        #PIG VARIABLES START
        export PIG_HOME=/usr/local/pig
        export PATH=$PATH:$PIG_HOME/bin
        #PIG VARIABLES END

-- Make sure the above syntax is identical with the spacing between the variables and =
-- Save and close the .bash_profile file.
-- Run the following command in the terminal: $source ~/.bash_profile

Start Pig:

Pig can run in 2 modes: MapReduce and local (standalone).

-- To start pig in the local mode, use command: $ pig -x local

-- To start pig in the MapReduce mode, use standalone command: $pig

NOTE: Sometimes Pig does not run in the MapReduce mode when Hive is installed. To get this working, comment the export command HADOOP_USER_CLASSPATH_FIRST=true in .bash_profile file (added when installing Hive, NA if Hive not installed). To run hive on the system, uncomment the export command again. Also, source the .bash_profile and restart Hadoop again.