Pig Installation Guide - 0.13.0

NOTE: If a dedicated Hadoop user (hduser) is present in Ubuntu, 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: $su hduser
-- Go to the home directory: $cd ~
-- Download Pig using the command: $wget 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 /home/user or similar
-- 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 (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 .bashrc 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): $su hduser
-- Go to the home directory inside hduser: $cd ~
-- Check if the .bashrc file exists: $ls -al
-- Open the .bashrc file using gedit/nano/vi: $sudo gedit .bashrc #Used gedit in this case
-- Append the following code to the end of the .bashrc 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 .bashrc file
-- run the following command in the terminal: $source ~/.bashrc

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 the .bashrc file (added when installing Hive, NA if Hive not installed). To run hive on the system, uncomment the export command again. Also, source the .bashrc and restart Hadoop again.