Pig Installation and execution of a query

1. Download pig from pig.apache.org

2. copy it to the hadoop user /home/hduser/

wget http://archive.apache.org/dist/pig/pig-0.13.0/[pig-0.13.0.tar.gz](http://www.us.apache.org/dist/pig/pig-0.13.0/pig-0.13.0.tar.gz)

tar xvzf pig-0.13.0.tar.gz

mv pig-0.13.0 pig

sudo mv /home/hduser/pig /usr/local/

**sudo chown -R hduser:hadoop /usr/local/pig**

**nano ~/.bashrc**

export PATH=$PATH:/usr/local/pig/bin

source ~/.bashrc

for (map reduce mode)

----------------------------

$ pig

or

for (local mode)

----------------------------

$ pig -x local

grunt>

----installation ends here ---

3. extract , make sure the extracted directory has the same permission (hduser:hadoop), insert path to $PATH

4. Start the hadoop server (./start-all.sh), check with jps

5. Copy the data to HDFS,

>hadoop dfs -copyFromLocal /user/hduser/<file-name>

6. run script using pig

>pig <script-name>

7. check if output is generated

>hadoop dfs -ls /user/hduser/<file-name>/

8. check output file

> hadoop dfs -cat /user/hduser/<output-dir>/part-r-00000