Doug was right, these are getting easier. Installing Pig is just like installing Hive.

Pig has two modes. They are specified by the -x option. If omitted, mapreduce is the default mode. Local is for running Pig on a single JVM for small projects, and for learning about Pig.

I clone my VM each week, so HDFS is already configured for a pseudo-distributed cluster. Pig mapreduce mode will run on either that or a fully-distributed cluster.

I used the -brief option in the last command to supress the timestamps.

[hadoop@pig ~]$ wget <http://apache.arvixe.com/pig/pig-0.15.0/pig-0.15.0.tar.gz>

[hadoop@pig ~]$ tar xzvf pig-0.15.0.tar.gz

[hadoop@pig ~]$ nano .bashrc

export PIG\_HOME=/home/hadoop/pig-0.15.0

export PATH=$PATH:$PIG\_HOME/bin

[hadoop@pig ~]$ source .bashrc

[hadoop@pig ~]$ pig -x local

grunt> quit

[hadoop@pig ~]$ start-dfs.sh

[hadoop@pig ~]$ jps

1616 NameNode

1714 DataNode

2012 Jps

1903 SecondaryNameNode

[hadoop@pig ~]$ pig -brief

grunt>